U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 403

TWO YEARS OF OPERATIONAL PREDICTIO OF FORECAST SKILL AT NMC

R. L. Wobus, and E. Kanlay
Development Division

October 1993

This is an unreviewed manuscript, primarily intended for informal exchange of information among NMC staff members

Two Years of Operational Prediction of Forecast Skill at NMC

Richard L. Wobus* and Eugenia Kalnay
Development Division
National Meteorological Center
Washington DC 20233

NMC Office Note 403

To be submitted to Monthly Weather Review

December 1993

*General Sciences Corporation, Laurel, MD

Abstract

At the National Meteorological Center we have been running in real time since 1990 a system to predict the forecast skill of the global spectral model, using as predictors the agreement of an ensemble of operational forecasts from various centers, the persistence in the forecast, and the amplitude of the anomalies. These are used in a multiple regression scheme with a 60-day training period, and we predict the regional anomaly correlation of the 00Z NMC global forecast from days 1 to 6. The most important predictor of skill is the agreement between the NMC global forecast started at 00Z, out to 6 days, and four other 12 hour "older" forecasts (JMA, UKMO, and ECMWF, as well as the average of the NMC forecast at 00Z with the previous day's forecast), so that this is like a "poor person's" Monte Carlo ensemble forecast. The other predictors have been selected to add to the predictive capability of the agreement alone, and together they quantify the factors that forecasters use subjectively when evaluating the available forecasts. These predictions are available to NMC forecasters on workstations and to outside users through Internet.

The predictive ability of this system, compares favorably with recent theoretical and experimental studies. The correlation between predicted and observed forecast skill seems to be best in regions where forecast skill varies significantly, and the seasonal variation in predicting the skill is small except in the tropics. The overall performance shows that these predictors include enough information about forecast skill to justify further development of skill predictions based on larger forecast ensembles and on more sophisticated statistical techniques.

1. Introduction

Since 1979, when the Global Weather Experiment took place, the skill of the operational forecasts has increased substantially in the short, medium and even longer ranges, both in the Northern and the Southern Hemispheres and in the tropics (e.g., Kalnay et al, 1990, Bengtsson, 1991). Nevertheless, forecast skill remains highly variable from day-to-day, region-to-region, and season-to-season. It is clear that the utility of the numerical forecasts would be considerably enhanced if, for example, a human forecaster could know that today's forecast will remain skillful for more than 6 days, or, conversely, that today's 3-day forecast will be much less reliable than normal. This fact prompted the often quoted statement of Tennekes et al. (1987) that "no forecast is complete without a forecast of the skill".

Epstein (1969) was the first to attempt to develop a method to predict the skill of dynamical forecasts. He introduced a stochastic-dynamic method for predicting the probability distribution of the model variables. This scheme, which requires prognostic equations for the covariances of the model variables, is too computationally expensive to be applied to realistic models. Epstein also suggested the use of ensemble forecasting for this purpose. Leith (1974) showed that in simulated "Monte Carlo Forecasting", when the perturbations were generated by random errors representative of the analysis uncertainty, even a small number of forecasts could improve the skill of the forecast mean and provide an estimate of its error. Hoffman and Kalnay (1983) suggested the use of the "Lagged Average Forecasting" (LAF) method, and showed that, at least for a simple model, LAF resulted in better predictions of the skill than the Monte Carlo method.

In principle, the theory of predictability provides the basis for prediction of the forecast skill. The fundamental theorem of predictability of Lorenz (1963) states that dynamically stable systems are infinitely predictable, whereas unstable systems have a finite limit of predictability. From this we can also conclude that the more unstable the atmosphere is, the less predictable it will be. In recent years there have been basically three approaches to the prediction of the forecast skill, all of which attempt to relate forecast skill to some measure of atmospheric stability:

a) Ensemble forecasting, which exploits the relationship between forecast agreement and forecast skill (e.g., Kalnay and Dalcher, 1987, Palmer and Tibaldi 1988, Murphy 1988, 1990, McCalla and Kalnay 1988, Kalnay and Ham, 1989, Tracton et al 1989, Baumhefner, 1991, Ebisuzaki and Kalnay 1991, Wobus and Kalnay 1991, Molteni and Palmer, 1993). In this approach it is assumed that if the model is sufficiently realistic, the same instabilities that increase the distance between a forecast and the real atmosphere, will also increase the "spread" among the members of the forecast ensemble. Two methods to generate perturbations for ensemble forecasting were recently implemented at NMC (Tracton and Kalnay, 1993, and Toth and Kalnay, 1993) and at ECMWF (Palmer et al 1992, Molteni and Palmer, 1993, Buizza, 1993), but are not yet used for prediction of the skill (see below).

b) Dependence of forecast skill on atmospheric regime, which is related to the variable stability properties of the large scale flow (e.g. Palmer 1988, Palmer and Tibaldi, 1988, Tracton et al, 1989, Tibaldi and Molteni, 1989, Molteni and Palmer, 1991). Similar to this approach are studies relating forecast skill to atmospheric persistence (e.g., Branstator 1986, Chen 1989), or to the Lorenz Index (Kimoto et al, 1991).

c) Regional prediction of maximum error growth based on the use of the adjoint of the forecast model (e.g., Barkmeijer, 1993, Houtekamer, 1993).

Barker (1991) performed a large number of Monte Carlo ensemble predictions using a simple but realistic primitive equations model (Roads, 1987). His results were rather discouraging: He used a large number of ensemble members, of the order of a hundred, and a "perfect model" simulation (i.e., one in which the same model was used to simulate the atmospheric evolution and the model forecasts). However, even under these favorable conditions, the correlation between the hemispheric forecast rms error, and the "rms spread" among the Monte Carlo ensemble members, varied between only 0.35 and 0.55 for the first 10 days of the forecast, and was even smaller later. This suggests that there is an upper limit for the predictability of the skill obtainable from ensemble forecasting. This is because, given a perfect model and a perfect knowledge of the statistical uncertainty in the initial conditions, the spread of ensemble trajectories can perfectly predict how fast another ensemble of trajectories (one of which could be the atmospheric trajectory) will drift apart. However, the ensemble spread cannot perfectly predict how a single forecast will compare with a single verification (H. van den Dool, 1992, pers. comm.). Moreover, Barker pointed out that when using Monte Carlo (random) perturbations, even if they represent perfectly the statistics of the simulated random analysis errors, the forecast of the skill has zero correlation between predicted and observed rms errors at the initial time. It is only after the growing perturbations organize themselves and dominate the error growth that the correlation between forecast agreement and forecast skill starts to increase. For this reason, his correlation between predicted and observed skill was 0.0 at the initial time and only 0.35 at day 1.

Palmer and Tibaldi (1988) and Molteni and Palmer (1991), have developed an experimental system to forecast the skill of the operational ECMWF forecasts. They use several predictors based on the ECMWF operational forecast, and predict the rms error directly and use statistical inference based on anomaly amplitude to predict the anomaly correlation.

More recently, Molteni and Palmer(1993), Mureau et al (1993), Buizza (1993), have developed a method to create dynamical perturbations for ensemble forecasting based on the adjoint method of Lorenz (1965). They chose the fastest growing linear perturbations for the first 36 hours of the forecast obtained as suggested by Lorenz. These perturbations are then combined to create the initial conditions of an ensemble of 31 forecasts each run at T63/19L model resolution. This method for ensemble forecasting was implemented operationally at ECMWF in December 1992, and is run 3 times per week in addition to the daily operational forecast that is run at T213/31L resolution. The advantage of this method is that the initial pertubations should be similar to the fastest growing modes that would also make the control operational forecast operational drift apart from the real atmosphere in the first day or two. In addition, the availability of a large number of forecasts should provide some quantitative basis for the issuing of probabilistic forecasts. The application of the ensemble prediction to forecast of the skill and to probability forecasting is still under development at ECMWF. Human forecasters in Europe currently have experimental access to the 31 individual forecasts as well as to clusters of forecasts.

At NMC, a different method to generate perturbations for ensemble forecasting, which should be representative of the fast growing errors present in the analysis cycle, was developed by Toth and Kalnay (1993). They use a very simple method denoted "breeding of growing modes" (BGM), in which the differences between short perturbed forecasts and the control forecasts are rescaled every 6 hours back to a fixed size similar to the analysis uncertainties, and then added as a perturbation to the initial condition of the next period's forecast. This process of repeatedly growing a perturbation and rescaling it, closely simulates the analysis cycle, in which a 6-hour forecast starting from an imperfect analysis grows away from the real atmosphere. The forecast is then used as a first guess for the next analysis, in which the use of data "scales down" the difference between the forecast and the atmosphere. Because the fastest growing errors dominate the growth in the short range forecast, the analysis cycle is a "breeding ground" for fast growing errors, which are damped but not eliminated by the data. The BGM method mimics this process, resulting in perturbations with characteristics similar to the growing errors present in the analysis. The BGM method is used every day at NMC to create perturbed forecasts at T62 resolution, extended to 12 days. This, combined with a control forecast and the extension of the 3-day "Aviation forecasts" results in the daily availability of 14 forecasts verifying for the same 10-day period. This method has been used to improve the average ensemble forecast, and to provide qualitative estimates of the outcome among different possible forecasts (Tracton and Kalnay, 1993). This ensemble forecasting system should also provide the basis for probabilistic forecasting at NMC.

The purpose of this paper is to review the results of a different method, also developed at NMC, for operational prediction of the regional forecast skill. This method, which has been extensively tested, is based on the use of forecasts from multiple centers, and other predictors of skill (Wobus and Kalnay, 1991). The NMC system for skill prediction has been running in real time since 1990, and has actually resulted in better correlations between predicted and observed skill than those obtained by Barker (1991). Its output is available in real time to any interested user through Anonymous File Transfer Protocol (FTP) in Internet (Appendix). While it is specifically designed to predict the skill of the NMC Medium Range Forecast model forecasts (MRF), the predictors it uses make it a generic prediction of the skill of various models. Section 2 contains a description of the method; Section 3 presents daily predictions of the skill for a recent season (spring of 1993); Section 4 contains comprehensive statistical verifications for all seasons and all regions of the world; and Section 5 a summary and discussion of future plans[1].

2. The NMC system for prediction of the skill

As indicated in the previous section, predictions of the skill based on ensembles have been generally based on either Monte Carlo Forecasting (MCF) or on Lagged Average Forecasting (LAF). For example, Dalcher et al (1987) tested the LAF method with a 100-day data set from ECMWF, and showed that the method showed promise in

---

[1] To keep the discussion as clear as possible, we use the term "forecast" to refer to a forecast of a future state of the atmosphere, as produced by a numerical model, and the term "prediction" to refer to a prediction of the skill of a forecast as measured by the anomaly correlation statistic.

predicting the skill. However, the decay of the forecast skill itself with time makes the use of LAF with once-a-day forecasts undesirable, because forecasts "older" by two or more days have inital errors much larger than those of the "younger" forecasts (Tracton et al, 1989).

To avoid this problem, as well as the need to perform additional LAF or MCF forecasts, we have developed a method in which the main predictor of skill is the average of the agreement between the NMC global forecast and other centers' operational forecasts (McCalla and Kalnay, 1988, Kalnay and Ham, 1989, Wobus and Kalnay, 1991). We take advantage of the fact that the NMC Medium Range Forecasts (MRF) are started daily at 00 UTC, whereas forecasts from three other global forecasting centers (United Kingdom Meteorological Office, Japan Meteorological Agency and European Centre for Medium-range Weather Forecasts) are started at 12 UTC. As a result, by the time the NMC MRF forecasts from 00 UTC become available to the users, forecasts from UKMO, JMA and ECMWF started 12 hours earlier are also available.

The NMC operational prediction of the skill is based on a multiple regression scheme using a 60-day training period. Because it is desirable to provide the forecaster with regional guidance rather hemispheric or global skill, we have chosen to predict the regional 500 hPa anomaly correlation of the 00 UTC NMC global forecast for regions of 30° latitude by 60° longitude covering the whole globe (Fig. 1). The most important predictor of skill is the agreement between the NMC medium range global forecast started at 00Z, and four other 12-hour "older" forecasts (JMA, UKMO, and ECMWF, as well as the average of the NMC forecast at 00 UTC with that of the previous day). Since the members of the ensemble of forecasts are obtained at essentially no additional cost, this can be considered a "poor person's" Monte Carlo ensemble forecasting scheme. On the other hand, it can be also considered as the most sophisticated and advanced ensemble that could be possibly created, since the ensemble members are made with state-of-the-art forecast models developed and improved over many years by many scientists, and the differences in their initial conditions truly reflect the uncertainties in our knowledge of the real atmosphere.

We produce predictions of the regional skill for the MRF forecasts at 12 UTC. Since the MRF is started at 00 UTC, the forecast of the skill is therefore valid for the 12 hr, 36 hr, 2.5 days,..., up to 5.5 forecast days. The upper limit of 5.5 days is forced by the fact that the UKMO forecasts are only 6-days long.

As mentioned before, the predictand is the regional 500 hPa anomaly correlation (AC) between the MRF forecast and the analysis, i.e., the regional pattern correlation between the forecast minus climatology and the analysis minus climatology. We have chosen the anomaly correlation as a measure of skill because it is the most widely used and easiest to interpret. The AC starts from an initial value of 1.0 (when the forecast is identical to the analysis) and, on the average, decays monotonically towards zero. The value of AC = 0.6, generally considered to be the minimum value for a forecast to retain useful skill, is attained on the average at about 6 to 8 days on a hemispheric basis, and somewhat earlier on a regional basis. On a daily basis, however, the variability of the AC is very large, with many cases of forecasts maintaining an AC of over 0.9 for over a week, or dropping below 0.0 in only 2 or 3 days (see, for example Figs. 5-7).

We are currently using three predictors of skill in the multiple regression scheme; all computed daily for each region and for each forecast length:

1) Forecast agreement (denoted as AGR), defined as the regional anomaly correlation between the MRF forecast and each of the other 12 hours "older" forecasts (UKMO, ECMWF and JMA, and the average of the latest two MRF forecasts). The individual AC between the MRF forecast and the other four forecasts are then averaged to create the forecast agreement predictor.

2) Forecast rms anomaly amplitude (RMS), defined as the regional rms amplitude of the MRF forecast anomaly with respect to climatology.

3) Forecast persistence (PERS), defined as the regional AC between the MRF forecast and the initial analysis.

Many other potential predictors of skill were also tested by Kalnay and Ham (1989), including regional values of baroclinic instability, Pacific North American (PNA) pattern index, regional values of the height and zonal and meridional wind values, etc., but they were shown to be less useful for the prediction of 0 to 6 day forecast skill than the three predictors chosen above. This does not mean that their correlation with forecast skill was lower than RMS and PERS, but that they resulted in less reduction of variance when combined with AGR. The stepwise regression procedure used in these tests and in the current skill prediction procedure attempts to combine predictors that contribute the most independent information to the regression.

Every day we develop prediction of skill equations for each region and forecast length by stepwise regression. We use as training data the same predictors and predictands computed from the forecasts corresponding to the previous 60 days. We do not constrain the regression equations to contain the same predictors from forecast day to forecast day, or from region to region. Thus, for each cycle, there are up to 60 forecasts available from dependent data, and one from independent data. It should be pointed out that the most important predictor, by far, is the forecast agreement, which is selected over 95% of the time, compared to RMS amplitude and forecast persistence, selected about 70% and 45% of the time respectively (Section 4). With respect to the length of the training period, experiments done by Kalnay and Ham (1989) indicated that the use of 60 days was better than using 30 days, and that the results were similar or better than those obtained using a longer 90 day training period. During our tests we have found that when 30 to 45 day training periods are used there are more frequent large errors in skill prediction associated with regime changes, particularly in the tropics. On the other hand, the advantage of using a recent series of forecasts is that the system quickly adapts to changes in the operational models or analysis systems employed as predictors.

The results of the multiple regression provide for each region and for all forecast lengths (from 0.5 days to 5.5 days) the predicted regional anomaly correlation. In addition, the system provides the average regional AC and the standard deviation of AC for the training period and the expected error of the predicted AC, obtained through the reduction of variance from the dependent sample. In addition, once it is available, the actual observed AC is provided.

A typical example of the regression equation, averaged for the spring of 1993, corresponding to region N11 (denoted "North America") and for the 3.5 day forecast is

$$AC_{pred}(N11, 3.5days) = 0.66 * AGR + 0.12 * RMS + 0.06 * PER + 0.08$$

In this case, the average value of the agreement was 0.83, the average RMS anomaly (in units of 100 m) was 0.98, and the average persistence was 0.29. Fig. 5a, discussed in the next section, shows the daily observed and predicted AC obtained for this period. It is clear from the coefficients in the equation that the agreement and the RMS amplitude dominate the contribution to the prediction of the anomaly correlation.

Figs. 2 a-f presents a sequence of predictions of the skill for the forecasts from 00 UTC 8 March 1993 through 13 March 1993, which included the verification of the "great blizzard of 1993" whose maximum amplitude was observed on 12 UTC March 13 on the east coast of North America. The forecast of the skill corresponds again to region N11 ("North America") in Fig. 1. Each figure shows with a thin line and vertical shading, the average AC observed in that region during the training period, as well as its standard deviation. Superimposed on it is the predicted AC and the observed AC (which is not available in real time). The forecaster can therefore observe how today's predicted forecast skill compares with the average skill of the model during the previous 60 days. In addition, the predicted skill plot also includes the expected error (dashed lines), based on the reduction of variance during the training period. This allows the forecaster to see the confidence that should be placed on the prediction of skill based on the training period. If the separation of the dashed lines is small compared with the width of the training period standard deviation, this implies a high degree of reduction of variance in the training period, and conversely, a skill prediction with a high standard deviation indicates that there is not much information in the training sample.

The sequence of figures shows that throughout the period 8-13 March 1993, the MRF forecast of the storm was generally excellent, and that this performance was also correctly predicted, enhancing the confidence in the forecast of this unusual event.

Figs. 3a and 3b present the 5.5 day forecast from 00 UTC 8 March 1993 and the corresponding verification valid on 12 UTC 13 March 1993, and Fig. 3c shows the error in the 500 hPa field at the verification time. Fig. 4 shows the forecast of the skill and its verification for region N9 ("Japan") from 00 UTC 8 March 1993, the same forecast cycle as in Figs. 2a and 3a. It is interesting to note that the actual errors in the "North America" and "Japan" regions are rather similar: they are both dominated by a rather large dipole of a high and a low oriented in the East-West direction, although the errors are out of phase. The amplitude of these dipoles is only slightly larger over "Japan" than over "North America", and in fact the RMS error is quite similar in both regions. Nevertheless, since the actual amplitude of the anomaly was huge over "North America", and relatively small over "Japan", the general direction of the flow was well predicted over "North America", but poorly predicted over "Japan". This is reflected in the prediction of skill over both regions (compare Figs. 4 and 2a), and both predictions of skill verified well in this case.

This example brings up an advantage of the use of AC as a measure of skill compared with RMS error. The AC is more of a relative measure of skill than the RMS error, and therefore is more useful in the case of large forecast anomalies, in which case the RMS error is likely to be large. The RMS error is a more absolute measure of skill, and therefore is more meaningful in the case of small forecast anomalies, when the AC tends to be poor because the signal-to-noise ratio is small. Since the forecaster has access to the actual forecast and its anomaly with respect to climatology, he or she can judge the significance of the predicted AC. This also indicates that the usefulness of RMS amplitude of the forecast anomaly as a predictor of AC is not simply a result of the fact that low RMS forecast amplitudes tend to be associated with low forecast AC due to the signal-to-noise effect, as suggested by Palmer and Tibaldi (1988). It also provides a useful quantification of the relative error measured by the AC, as in the above example.

The results of the forecast of the skill are available to the NMC forecasters as shown above in graphical form, but as indicated before, they are also available in digital form through the use of anonymous FTP within Internet (see the Appendix for a documentation of how to access the NMC public file server in real time).

The example presented above is an exceptionally good example of predictions of the skill, associated with an excellent forecast over North America (and a poor forecast over eastern Asia). In the next section we compare the predicted and observed AC for a complete season, for a few selected regions, and in Section 4 we present comprehensive statistics for all regions, forecast lengths and seasons.

3. Daily results for the spring of 1993

We have shown in Section 2 an example of how the prediction of the regional forecast anomaly correlation is presented to the forecaster on a workstation or personal computer. In this section we present a comparison between 90 days of daily predicted and observed forecast AC for several representative mid-latitude Northern Hemisphere regions, as well as for one extratropical region in the Southern Hemisphere and one tropical region. As will be shown in the next section, the high latitude (60°-90°) statistics are very similar to those of the middle latitudes (30°-60°), so that in the interest of space we do not present high latitude examples. In each of the following figures the correlation ρ between predicted and observed AC for the whole season is also indicated.

Figs. 5 a-c presents the predicted and observed regional AC for the 3.5 day forecasts over regions N11, N7, and N9 ("North America", "Europe" and "Japan" respectively) verifying during March, April and May of 1993. The prediction of skill over North America is excellent (ρ=0.73), with most of the major maxima and minima in forecast skill well predicted. It should be noted that the scheme seems to predict quite well almost all the extreme events during this period, which is remarkable for a regression algorithm. Over "Europe", where ρ=0.48, the variability in skill is much lower than over "North America," and the forecast of the skill is able to capture the low frequency variability in skill, but not much of the the day-to-day variation. Over "Japan" ρ=0.72, and the prediction of skill, like over "North America," captures most of the major maxima and minima, but misses some of the daily variability. The seasonal trend has not been subtracted from the AC's, but these figures (and those for the other seasons), do not show an appreciable effect of the seasonal cycle.

Figs. 6 a-c present the same results but for the 5.5 day forecast. In this case the correlation ρ over "Europe" is only 0.34, and in this area the skill prediction scheme is not useful on a day-to-day basis, although it does seem to capture the intraseasonal variability in skill. Over North America, the correlation is 0.45, and although the scheme captures some of the low frequency variability in skill, it generally underestimates it. A notable exception is the 6-day period around March 10, when the scheme consistently and correctly predicted high skill associated with the forecast of the East Coast blizzard of 93. The 5.5 day prediction over Japan, on the other hand, continues to be quite skillful, with the scheme still capturing both high and low frequency variability, and a correlation ρ of 0.61.

Finally, Figs. 7 a-b show the 3.5 day prediction of skill for the mid-latitude region S11 ("Southern Cone") and for the tropical region S15 ("Australia"). For the "Southern Cone" ρ=0.61, and the prediction is quite good, even on a daily basis. For "Australia," ρ=0.61, and most of the skill is attained for the low frequency variations, as is the case in general for the tropics.

A subjective evaluation of many time series like these suggest that the prediction of skill should be useful, at least for the low frequency variability in skill, if the correlation between predicted and observed anomaly correlation is above 0.4. If it is above 0.6, then it is probably useful even in the prediction of day-to-day variability in skill.

4. Statistical verifications for all seasons, regions and forecast lengths

In this section we present verification statistics for the two years (8 seasons) for which the skill prediction system has been operationally available. It should be noted that the Analysis/Forecast system used for the Medium Range Forecast (MRF) model at NMC remained relatively stable during this period, which spans June 1991 through May 1993. Although a number of relatively minor improvements were implemented, the only major change was the replacement of the Optimal Interpolation (OI) analysis by the Spectral Statistical Interpolation (SSI, Parrish and Derber, 1992, Derber et al, 1991), a 3 dimensional variational analysis system, which took place on June 25 1991. This change, which affected most dramatically the tropics by eliminating or drastically reducing the model spinup, influenced the results for the summer season (JJA 1991) both directly and through the 60-day training period.

a) Average statistics for one year.

We now present a number of statistical characteristics of the skill prediction system averaged for all latitudinal bands in both hemispheres, and averaged over the last four recent seasons (summer of 1992 through spring of 1993). It should be noted that the annual average is obtained as the average of four individual seasonal averages, so that the correlation ρ between observed and predicted AC's is not significantly affected by the seasonal cycle.

In fig. 8 we present the observed Northern Hemisphere regional anomaly correlations averaged for the four seasons, and for the latitude bands. It is apparent that the forecast AC of the NMC medium range model computed regionally remains above 60% till about day 6, even on the annual average. The skill for mid- and high-latitudes is virtually identical, and displays the S-shaped, "convex up" curvature which is characteristic of AC of forecasts whose main source of errors is the unstable growth of initial errors (Reynolds et al., 1994), and for which errors due to model deficiencies are relatively very small. On the other hand, the AC over the tropics decreases initially much faster than in the extratropics, although there seems to be a crossover point at about day 6. The AC for the tropics exhibits the "concave up" shape that according to Reynolds et al. (1994) is characteristic of forecasts whose errors are dominated by model deficiencies, and not by atmospheric instabilities.

Fig. 9 is like Fig. 8 but showing the forecast agreement instead of the anomaly correlation. It bears a remarkable resemblance to Fig. 8. A careful comparison between the two figures shows that the forecast agreement over mid- and high latitudes is only marginally higher than the forecast anomaly correlation with the real atmosphere, whereas for the tropics, the forecast agreement is actually lower than the AC. The similarity in shape between the two figures shows that the divergence among forecasts is dominated also by instabilities in the extratropics and by differences among the model behaviour, presumably due to different physical parameterizations, in the tropics. This figure shows that if we compare the NMC model with an ensemble of 12-hour "older" operational models, the often repeated statement that "any two forecasts resemble each other much more than any one of them resembles the real atmosphere" is not true any more with present state-of-the-art systems. This realistic divergence among forecasts coming from different centers is probably one of the most important reasons for the relative success of the present method for predicting skill, as will be discussed later.

Fig. 10a presents the correlation ρ between predicted and observed AC for the dependent (training) sample. It is somewhat surprising that the ρ is significantly higher for the tropical regions than for the extratropics, reaching a peak of 0.75 at day 2.5 in the tropics. For mid- and high-latitudes, the dependent sample correlation decreases slowly from about 0.7 at 0.5 days, to less than 0.6 at day 5.5. Fig. 10b shows the same correlation ρ for both the dependent and the independent (actual forecasts) samples. As could be expected, the scores for the independent samples are lower than those for the dependent samples, by about 10 to 15%. The mid-latitudes skill in predicting the AC decreases from 0.65 at day 1.5 to 0.45 at day 5.5. The high latitudes scores are about 5% lower. Once again, the tropics are the regions where the forecast of the skill is best. As discussed later in this section, the higher scores in forecasting the skill over the tropics is probably due to the fact that the anomaly correlations themselves are much more variable and have much more low frequency variability in the tropics than in mid- or high-latitudes, and therefore there is "more room" to capture variability in the signal.

In Fig. 10c, we show once again the correlation ρ for the actual operational forecasts of the skill, but this time we also include a similar correlation obtained by Barker (1991) using a perfect model Monte Carlo ensemble rms spread to predict the rms forecast error. It is remarkable that our operational results are actually better than those of the perfect model MC experiments of Barker, who used a much larger number of ensemble members (120 versus only 4 in the NMC system), and who was not encumbered by model deficiencies in his simulation. The fact that at short forecast intervals the MC results are poorer than the NMC results should not be surprising, since, as mentioned before, a pure MC approach starts from random pertubations, and therefore has zero predictive skill at the initial time. The NMC operational system, on the other hand, starts from very "realistic" perturbations, since each of the operational systems is a state-of-the-art system, and each one of their analysis cycles is a "breeding ground" for the same type of fast growing errors that plague all operational forecasts (Toth and Kalnay, 1993). Beyond day 3, the results are more similar between the MC and the operational system, although the latter remains clearly superior. The difference between

Barker's results and those of NMC is important since, as indicated before, skill predictions with correlations above 0.4 seem to have some usefulness, especially for low frequency variability, and those above 0.6 appear to be quite useful.

There are several additional possible explanations for such superiority: The NMC system predicts AC, and Barker (1991) predicts rms error. The NMC system uses other predictors, in addition to forecast agreement, i.e., forecast persistence, a proxy for atmospheric stability, and the rms amplitude of the anomaly. The last predictor is especially useful for predicting AC (Branstator, 1986), since at low values of the anomaly the signal-to-noise ratio of the AC becomes small (Palmer and Tibaldi, 1988). Nevertheless, as shown in the example of the prediction of skill from the forecast from March 8 1993, we cannot conclude that the advantage of higher predictability that the AC has when compared to rms error, is simply a statistical artifact: the rms error over "Japan" and "North America" were rather similar in both shape and size, but the anomalous circulation was far stronger over the U.S. Therefore, an rms error score would have indicated similarly poor forecasts, whereas the AC score correctly indicated that the anomalous circulation was not well captured over Japan, and very well represented over North America.

In addition, it is possible that the "poor person" Monte Carlo system adopted at NMC, having systems with many more degrees of freedom than the simpler model of Barker (1991), has more skill variability, and therefore "more room" to predict the skill. As mentioned above, we see this effect in Fig 7b ("Australia", a tropical region), when compared to the other mid-latitude regions of Figs. 5 and 7a. This may also explain the difference with the results of Houtekamer and Derome (1993), who used a "breeding cycle" to generate analysis errors as well as ensemble perturbations, instead of starting with random MC perturbations. When they performed perfect model ensemble forecasts and used, as Barker, the ensemble spread to forecast the rms errors, they obtained correlations between 0.4 and 0.47 throughout the first 10 days of the forecast. This relatively low value for a perfect model system may be due to the use of a T21 quasi-geostrophic model, with fewer degrees of instability than the real atmosphere or the operational models.

In Fig. 11 we compare the actual correlations ρ between predicted and observed AC for the Southern Hemisphere with those of the Northern Hemisphere. The scores of the Southern Hemisphere are about 10% worse for both the tropics and the extratropics, possibly due to a lower general of skill of the forecasts themselves, associated with poorer observations and therefore poorer initial conditions.

Fig. 12a compares regional correlations ρ for several regions of the Northern Hemisphere, and Fig. 12b, in the Southern Hemisphere. "North Africa", an example of a tropical region, has a correlation between predicted and observed AC varying between 0.6 and 0.73, "North America" and "Japan" have values of ρ between 0.5 and 0.7, and "Europe" starts at 0.62 at day 0.5, but the skill decreases to only 0.3 at day 5.5. The low correlation over Europe, and the high correlation over North Africa are probably associated with a much better overall forecast skill over Europe than over Africa, and therefore with much smaller forecast skill variability. The comparison is similar over the Southern Hemisphere (Fig. 12b), where the three tropical regions "Brazil", "Australia" and "Peru" maintain a correlation close to 0.6 up to day 3.5, whereas the mid-latitudes "Southern Cone" has a correlation between 0.45 and 0.55.

Fig. 13 presents the annual average of the percentage of cases in which the three predictors, forecast agreement, forecast persistence and rms amplitude of the anomaly, are selected as predictors for the multiple regression during the training. It shows that agreement is selected between 90% and 100% of the predictions, rms amplitude between 65% and 80%, and forecast persistence in 40% to 50% of the predictions. These percentages are essentially independent of the forecast length and are also similar for tropical and extratropical regions. Recall that after the first predictor is chosen, additional predictors are selected based on their contribution of additional information in the regression.

c) Variability over eight seasons.

Next we present some statistics indicating the variability of the statistics presented above with season. The available data corresponds to two years starting with the summer (JJA) of 1991, and ending with the spring (MAM) of 1993. For the sake of brevity, results are only presented for the forecast length of 3.5 days and for the Northern Hemisphere.

Fig. 14 shows both the AC and the correlation ρ for all eight seasons and for both mid-latitudes and the tropics. For the mid-latitudes the seasonal dependence is weak, with the average AC at day 3.5 over 0.85, except during the summers, when it is slightly lower. The correlation ρ between predicted and observed AC for mid-latitudes varies between 0.5 and 0.6, and shows even less seasonal dependence. The AC in the tropics, like the mid-latitudes, has the poorest forecast skill occurs during the summers, with higher forecast skill during the other seasons, during which intrusions of mid-latitude air masses make the tropical flow more predictable (G. White, pers. comm.). In the tropics it is clear that there is an upward trend in skill with time during these two years, with an AC of almost 0.8 for the last 3 seasons, presumably associated with the implementation of the SSI (which took place during the first summer season) and other less important improvements. The worst correlation ρ for the tropics occurred also during the first season, which was affected not only by the implementation of the SSI at the end of June 1991, but also but the inhomogeneity of the training period of 60 days previous to each forecast. The second summer is also relatively poor in the tropics.

Figs. 15 a-c show for several regions, the actual average AC, the correlation ρ between predicted and observed AC for both the training period (dependent sample) and the actual predictions of skill (independent sample). Once again, the individual mid-latitude regions ("North America," "Europe" and "Japan") do not exhibit a clear seasonal dependence of AC, except for lower values in the summer. "Europe" shows a dramatic improvement of AC after the implementation of the SSI. "North Africa," like the previous tropical averages, shows the lowest skill in the summer, and a clear upward trend in forecast skill.

The correlation ρ in the same four regions has more variability but does not show a seasonal trend. On the other hand, it is very encouraging that there is a certain similarity in the variation over the 8 seasons between the correlation ρ obtained for the training sample (which is the basis for the error bars of the predicted AC in Fig. 5) and the actual observed correlation ρ. This suggests that, to some extent, the average accuracy of the forecast of the skill can be estimated from the dependent sample.

5. Discussion

We have presented the results of a real time scheme to predict the global model forecast skill which has been operational at NMC for over two years, and is also electronically available through Internet (Appendix). The period for which we have shown results has been relatively homogeneous for the NMC model, and includes a major improvement in the analysis (the SSI implemented in June 1991, which affected mostly the tropics, and introduced problems with the training). We have presented the results for the daily prediction of skill for the last available season, and comprehensive summary regional statistics for all the regions and latitude bands, as well as their seasonal dependence over the two years.

We find that for most mid-latitude regions, and most seasons, the correlation between observed and predicted skill at 3-4 days is between 0.4 and 0.7, whereas over the tropics it is even higher. An inspection of the daily plots indicates that the predictions of skill should be useful in if the correlation between predicted and observed AC is above 0.4, especially in the prediction of the low frequency variability in the skill. When the correlation is 0.6 or higher we find that there is significant skill in predicting the day-to-day variability in forecast AC. A 5-day running mean of the predicted and observed AC results in a considerable improvement of the correlations, typically of 0.1 to 0.2, confirming that there is considerable predictability of low frequency variability in the skill (Kalnay and Ham, 1989).

Our results are somewhat better than those obtained by Barker (1991) and by Houtekamer and Derome (1993) using perfect models and large ensembles. They obtained correlations between predicted and observed skill of only about 0.3-0.4. Although our results are less discouraging than theirs, they also confirm that the prediction of skill is a very difficult problem. This is due to the fact that the actual trajectory followed by the atmosphere is only one of an ensemble of possible trajectories. It would be possible to perfectly predict the skill of a large ensemble of forecasts if we compared them against an ensemble of possible verifications. In practice, however, the atmosphere goes through a single realization, so that the prediction of the skill of a single individual forecast or even of an ensemble of forecasts is much less than perfect, as shown in the experiments of Barker (1991). This also suggest that in the future we should develop a system to predict the realiability of probability forecasts, a task made possible by the implementation of ensemble forecasting at NMC (Tracton and Kalnay, 1993).

The results of our system can be considered as moderately encouraging, even though we are using an extremely small ensemble (4 members in addition to the base NMC forecast) compared to Barker (1991) and Houtekamer and Derome (1993) who used hundreds of ensemble members. We attribute this to the fact that we are using one of the most realistic ensembles possible: The members are each an operational forecast derived independently in a different global forecasting center, with initial conditions also derived from independent analysis cycles. Both the analysis cycles and the models are somewhat different from each other, and, given the friendly competition among centers to show good forecast skill, they represent each a slightly different but most advanced state-of-the-art system. Therefore, the differences among the analyses and among the forecast models represent in the most realistic way the present uncertainties about our knowledge of the state of the atmosphere and the detailed laws that govern its evolution. For this reason, for example, our forecast of the skill starts and retains a correlation between predicted and observed AC of over 0.6 for several days, whereas Barker (1991), using a perfect model and a very large ensemble of forecasts (but starting with Monte Carlo perturbations) had to start with zero correlation at the initial time.

In addition to the use of very realistic ensemble members, we are helped by the use of the anomaly correlation as a measure of skill, rather than the rms error. This also allows the use of an additional powerful predictor of skill, the anomaly amplitude. As discussed in Section 2.3 this additional information is not just a simple exploitation of the "signal-to-noise-ratio" which occurs both in the forecast agreement and in the actual forecasts AC. Although it is true that very small anomalies are associated with low AC as well as with low agreement (Palmer and Tibaldi, 1988), this is only one advantage of AC. As discussed in the example of the case of 8 March 1993, the rms errors over the "Japan" and "North America" regions were quite similar in shape and magnitude. Nevertheless, the anomaly itself was huge over North America (the "great Blizzard of 1993"), and relatively small over Japan. Therefore, the forecast over "North America" captured well the observed anomalous circulation, whereas over "Japan" the forecast of the regional circulation was very poor. Therefore, even though the absolute regional rms errors were quite similar, the forecast AC was unusually high over "North America" and very low over "Japan." The forecast of the AC captured correctly this difference, undoubtedly taking advantage of the additional information provided by the forecast of the anomaly amplitude (Branstator, 1986).

The system described here has used the same ensemble members since the JMA forecasts became available at NMC in 1991. We plan to add more members to the ensemble as they become available to us in real time via global telecommunications, such as those from Canada, Australia, Germany, and the US Navy, which should improve the short range prediction of skill.

We are considering testing Kalman filtering and/or neural networks as an alternative to the linear regression method used here. These methods would eliminate the need for a hard cutoff to the training period, and would probably simultaneously improve the timeliness of the prediction model and the response to regime transitions. The results presented here indicate that there is enough information in our predictors to use either of these methods effectively. We are also currently developing new predictors based on the breeding ensemble (Toth and Kalnay, 1993), to be used in a manner analogous to the predictors we now use, but for forecast ranges to 10 days, or, if the ensemble forecasts are extended to a month, provide forecast of the skill for appropriate time averages. Finally, we plan to develop a system to predict the reliability of probability forecasts based on the NMC ensemble forecasting system.

Appendix: Access to operational prediction of MRF skill through Internet

The following 8 datasets are available from NMC through Internet in the NMC public access files:

```
     wd20rw.skill.mrf0        skill forecasts for today's MRF
     wd20rw.skill.mrf1        skill forecasts for yesterday's MRF
     ....
     wd20rw.skill.mrf7        skill forecasts for MRF 7 days ago
```

These datasets are regenerated daily to show the most recent verifications. The update takes place normally between 09Z and 11Z, but may be as late as 18Z.

Typical commands to download these datasets using anonymous ftp:

```
     type ftp command:   ftp sun1.wwb.noaa.gov<cr>
     type username:      anonymous<cr>
     type password:      _your_name_or_email_address_<cr>
     change directory:   cd /pub
     type get command:   get wd20rw.skill.mrf0 yourfn<cr>
     type bye command:   bye<cr>
```

where yourfn is the name of the file on your computer into which to download the skill data.

Format of these datasets:

Each dataset consists of 218 lines of 25 characters each.

Each line consists of 5 numbers, each occupying 5 columns.

Lines 1 and 110 consist of year (2 digits), month, day, hour (always 0), and 0.

Lines 2 through 109 (for the northern hemisphere) and 111 through 218 (for the southern hemisphere) consist of 6 lines (12 hr, 36hr, 60 hr, 84hr, 108hr, 132hr predictions) for each region (see back). The format is as follows:

tac tsd pac psd vac

in which:

tac=average training anomaly correlation
tsd=standard deviation of training anomaly correlation
pac=predicted anomaly correlation
psd=standard deviation of predicted anomaly correlation
vac=verifying anomaly correlation

These integers should be converted to floating point and divided by 1000.

The number -9999 represents a missing value, normally a verifying anomaly correlation for a future time.

6. References

Barker, T. W., 1991: The relationship between spread and forecast error in extended range forecasts. *J. Climate*, 4, 733-742.

Barkmeijer, J., 1992: Local error growth in a barotropic model, *Tellus*, **44A**, 314-323.

Barkmeijer, J., 1993: Local skill prediction for the ECMWF model using adjoint techniques. *Mon. Wea. Rev.*, **121**, 1262-1268.

Baumhefner, D. P., 1991: Evaluation of long-range forecast skill from Monte Carlo ensemble integrations of a low resolution NCAR Community Climate Model. Proc. of the NATO Workshop on Prediction of Interannual Climate Variations, J. Shukla, Director. 22-26 July 1991, ISAS, Trieste, ITALY. NATO ASI Series Publication, Springer-Verlag.

Bengtsson, L. 1991: Advances and prospects in numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **117**, 855-902.

Branstator, G., 1986: The variability in skill of 72-hour global-scale NMC forecasts. *Mon. Wea. Rev.*, **114**, 2628-2639.

Buizza, R. 1993: Unstable perturbations computed using the adjoint technique. Submitted to *Quart. J. Roy. Met. Soc.*

Chen, W., 1989: A new method in forecasting forecast skill. *Mon. Wea. Rev.*, **117**, 427-435.

Dalcher, A., E. Kalnay and R. Hoffman, 1988: Medium-range lagged averaged forecasts. *Mon. Wea. Rev.*, **116**, 402-416.

Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. and Forecasting*, 6, 538-547.

Ebisuzaki, W., and E. Kalnay, 1991: Ensemble experiments with a new Lagged Analysis Forecasting scheme. Research Activities in Atmospheric and Oceanic Modelling, Report No. 15, WMO.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739-759.

Hoffmann, R. and E. Kalnay, 1983: Lagged average forecasting: an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100-118.

Houtekamer, P., 1993: Global and local skill forecasts. *Mon. Wea. Rev.*, **121**, 1834-1846.

Houtekamer, P. and J. Derome, 1993: Prediction Experiments with two-member ensembles. Submitted to *Mon. Wea. Rev.*

Kalnay, E. and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.*, **115**, 349-356.

Kalnay, E., and M. Ham, 1989: Forecasting forecast skill in the Southern Hemisphere. Preprints of the Third International Conference on Southern Hemisphere Meteorology and Oceanography. Buenos Aires, 13-17 November 1989. AMS publisher.

Kalnay, E., M. Kanamitsu, and W. E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Am. Meteor. Soc.*, **71**, 1410-1428.

Kimoto, M., H. Mukougawa and S. Yoden, 1992: Medium-range forecast skill variation and blocking transition: a case study. *Mon. Wea. Rev.*, **120**, 1616-1627.

Leith, C. E, 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.*, 20, 130-141.

Lorenz, E. N., 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus*, 17, 321-333.

McCalla, C. and E. Kalnay, 1988: Short and medium forecast skill and the agreement between operational forecasts. Preprints of the 8th AMS Conference on Numerical Weather Prediction. Baltimore, MD, February 1988, 634-640.

Molteni, F., and T. N. Palmer, 1991: A real-time scheme for the prediction of forecast skill. *Mon. Wea. Rev.*, **119**, 1088-1097.

Molteni, F. and T. N. Palmer, 1993: Predictability and finite-time instability of the Northern Winter circulation. *Quart. J. Roy. Met. Soc*, **119**, 269-298.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Met. Soc*, **119**, 299-323.

Murphy, J., 1988: Impact of ensemble forecasts on predictability. *Quart. J. Roy. Met. Soc.*, **114**, 463-493.

Murphy, J., 1990: Assessment of the practical utility of extended-range ensemble forecasts. *Quart. J. Roy. Met. Soc.*, **116**, 89-125.

Palmer, T. N., 1988: Medium and extended range predictability and stability of the PNA mode. *Quart. J. Roy. Met. Soc.*, **114**, 691-713.

Palmer, T. N., and S. Tibaldi, 1988: On the prediction of forecast skill. *Mon. Wea. Rev.*, **116**, 2453-2480.

Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet and J. Tribbia, 1992: Ensemble Prediction. ECMWF Research Department Tech. Memo. No 188. Available from ECMWF, Reading, UK.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **128**, 1747-1763.

Reynolds, C., P. J. Webster, and E. Kalnay, 1994: Random error growth in the NMC Medium Range Forecast Model. *Mon. Wea. Rev.*, accepted.

Roads, J. O. (1987): Predictability in the extended Range. *J. Atmos. Sci.*, 44, 3495-3527.

Tennekes, H. A., P. M. Baede, and J. D. Opsteegh, 1987: Forecasting forecast skill. Proc. ECMWF Workshop on Predictability in the Medium and Extended Range. Reading, UK. Available from the European Center for Medium range Weather Forecasts.

Tibaldi, S. and F. Molteni, 1990: On the operational predictability of blocking. *Tellus*, **42A**, 343-365.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Met. Soc.*, in press.

Tracton, S., K.-C. Mo, W. Chen, E. Kalnay, R. Kistler and G. White, 1989: Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.*, **117**, 1604-1635.

Tracton, S., and E. Kalnay, 1993: Operational ensemble forecasting at NMC: Practical aspects. *Wea. Forecasting*, 8, 379-398.

Wobus, R. L., and E. Kalnay, 1991: Prediction of forecast skill for the NMC global model. Preprints of the 9th AMS Conference on Numerical Weather Prediction. Denver, CO, October 14-18 1991, 481-484.

Figures:

Fig. 1. Locations of the regions for which the skill prediction is performed.

Fig. 2. Skill predictions with verification for region N11 ("North America") for MRF forecasts from 00Z of initial dates: (a) 8 March 93, (b) 9 March 93, (c) 10 March 93, (d) 11 March 93, (e) 12 March 93, (f) 13 March 93. The "Great Blizzard of 1993" had its maximum amplitude on the East Coast on March 13 1993. In each figure, the shaded area corresponds to the training period, the thin lines are the forecast of the AC and its expected error, and the thicker line is the actual AC of the MRF forecast.

Fig. 3. 500 hPa MRF 5.5 day forecast (a), verification analysis (b), and error (c), valid 12Z 13 March 1993.

Fig. 4. Skill prediction with verification for region N9 ("Japan"), for the same forecast cycle as Fig. 2a.

Fig. 5. Predicted (solid) and verifying (dashed) anomaly correlation for 3.5 day MRF forecasts for (a) region N11 ("North America"), (b) region N7 ("Europe"), (c) region N9 ("Japan") verifying during March, April, and May 1993.

Fig. 6. Predicted (solid) and verifying (dashed) anomaly correlation for 5.5 day MRF forecasts for (a) region N11 ("North America"), (b) region N7 ("Europe"), (c) region N9 ("Japan") verifying during March, April, and May 1993.

Fig. 7. Predicted (solid) and verifying (dashed) anomaly correlation for 3.5 day MRF forecasts for (a) region S11 ("Southern Cone"), (b) region S15 ("Australia"), verifying during March, April, and May 1993.

Fig. 8. Anomaly correlation of MRF forecasts, annual average, by northern hemisphere latitude band.

Fig. 9. Agreement between MRF and other forecasts, ensemble AC, annual average, by northern hemisphere latitude band.

Fig. 10. Correlation between predicted and observed AC of MRF forecasts (a) for dependent sample, (b) comparing dependent and independent samples, (c) comparing independent sample with perfect model MC forecast (Barker, 1991), annual average, by northern hemisphere latitude band.

Fig. 11. Correlation between predicted and observed AC of MRF forecasts for independent sample, annual average, by latitude band in both hemispheres.

Fig. 12. Correlation between predicted and observed AC of MRF forecasts for independent sample, annual average by region for (a) northern hemisphere and (b) southern hemisphere.

Fig. 13. Frequency of selection of predictors, annual average, by northern hemisphere latitude band. The lines are labeled by predictor but not by latitude band.

Fig. 14. Correlation between predicted and observed AC of MRF forecasts for independent sample, by season and latitude band.

Fig. 15. Correlation between predicted and observed AC of MRF forecasts for dependent (training) and independent sample, with observed anomaly correlation, by season, for regions (a) N11 ("North America"), (b) N7 ("Europe"), (c) N9 ("Japan"), and (d) N1 ("North Africa").

Fig. 1

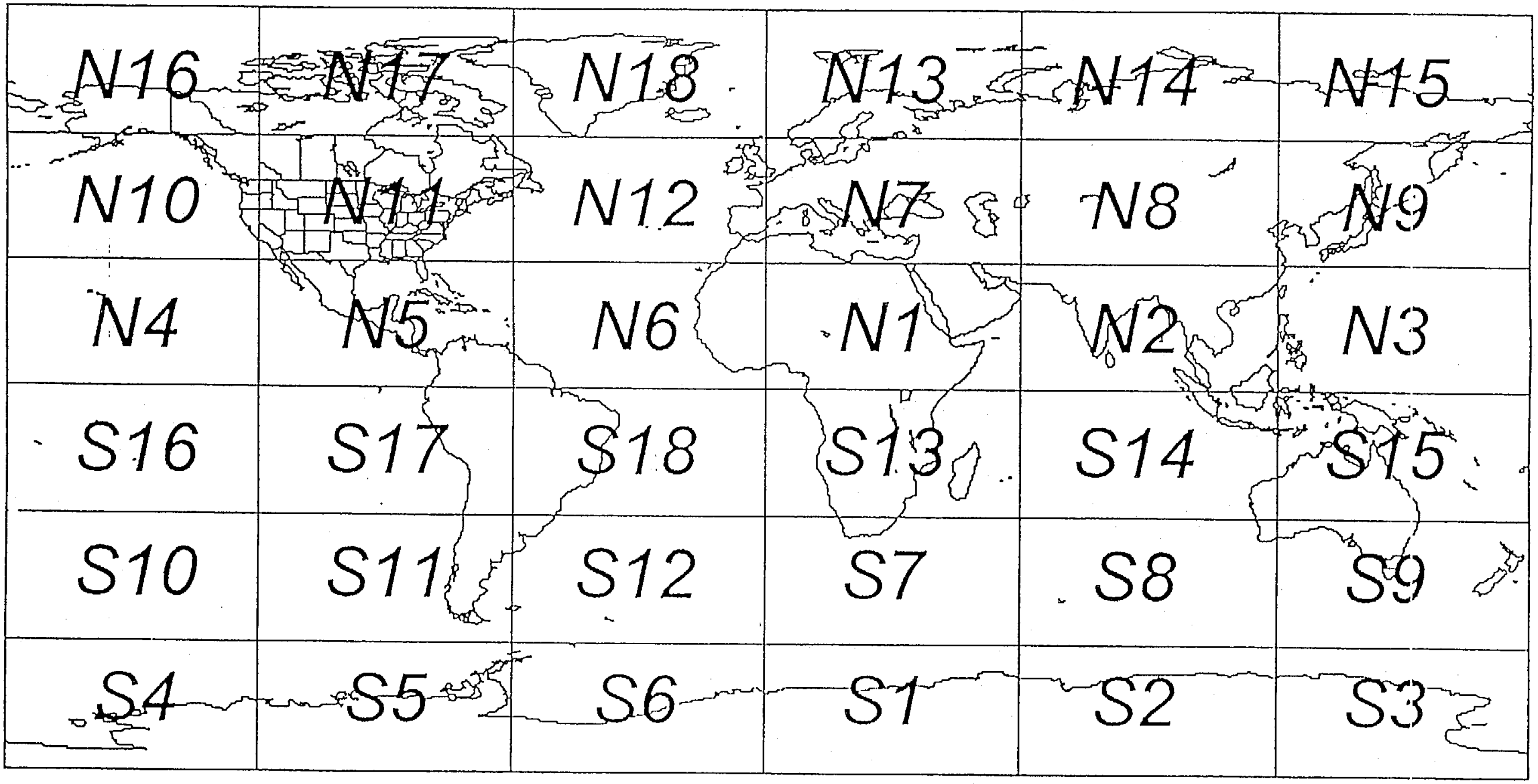

N16
N17
N18
N13
N14
N15
N10
N11
N12
N7
N8
N9
N4
N5
N6
N1
N2
N3
S16
S17
S18
S13
S14
S15
S10
S11
S12
S7
S8
S9
S4
S5
S6
S1
S2
S3

Fig. 1

Fig. 2a

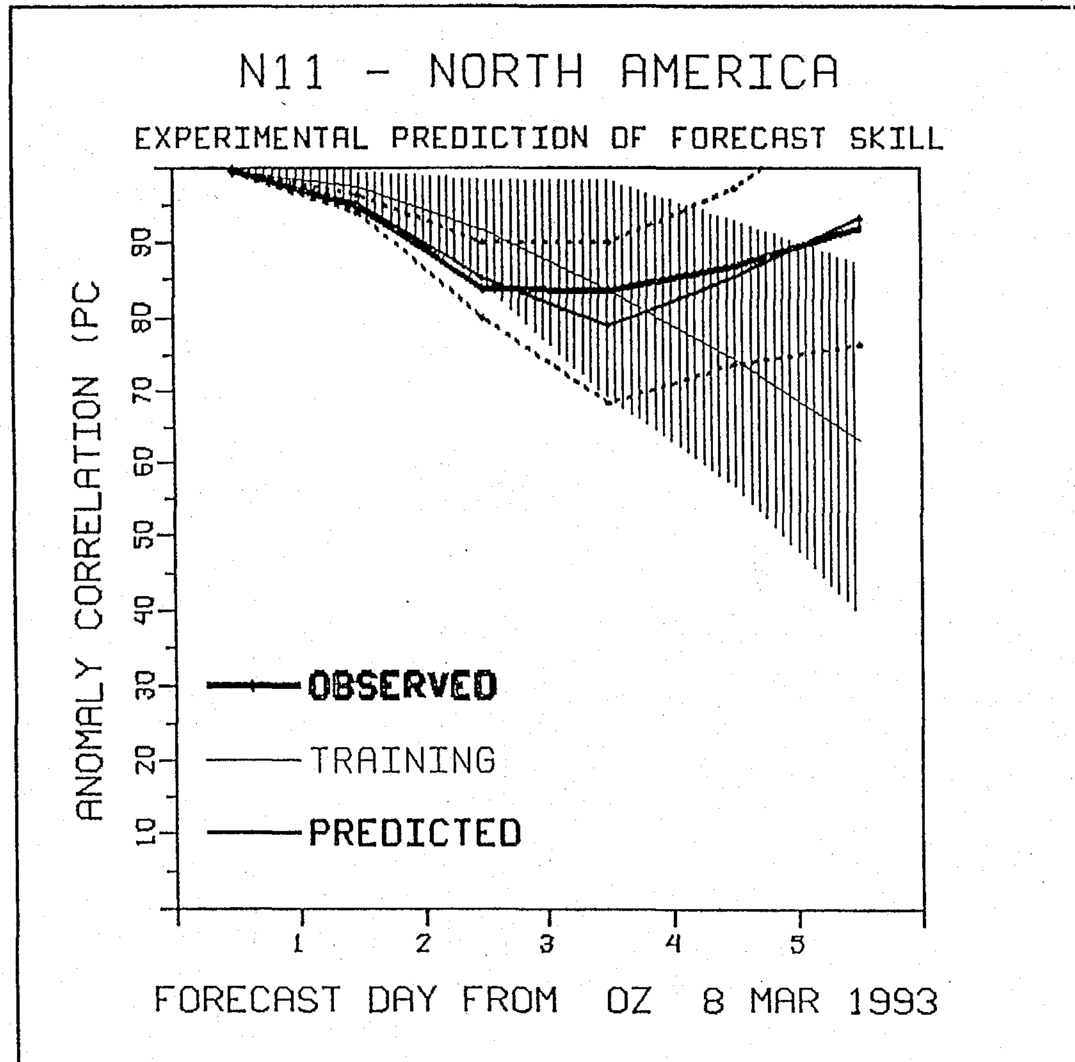

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM 0Z 8 MAR 1993

Fig. 2a

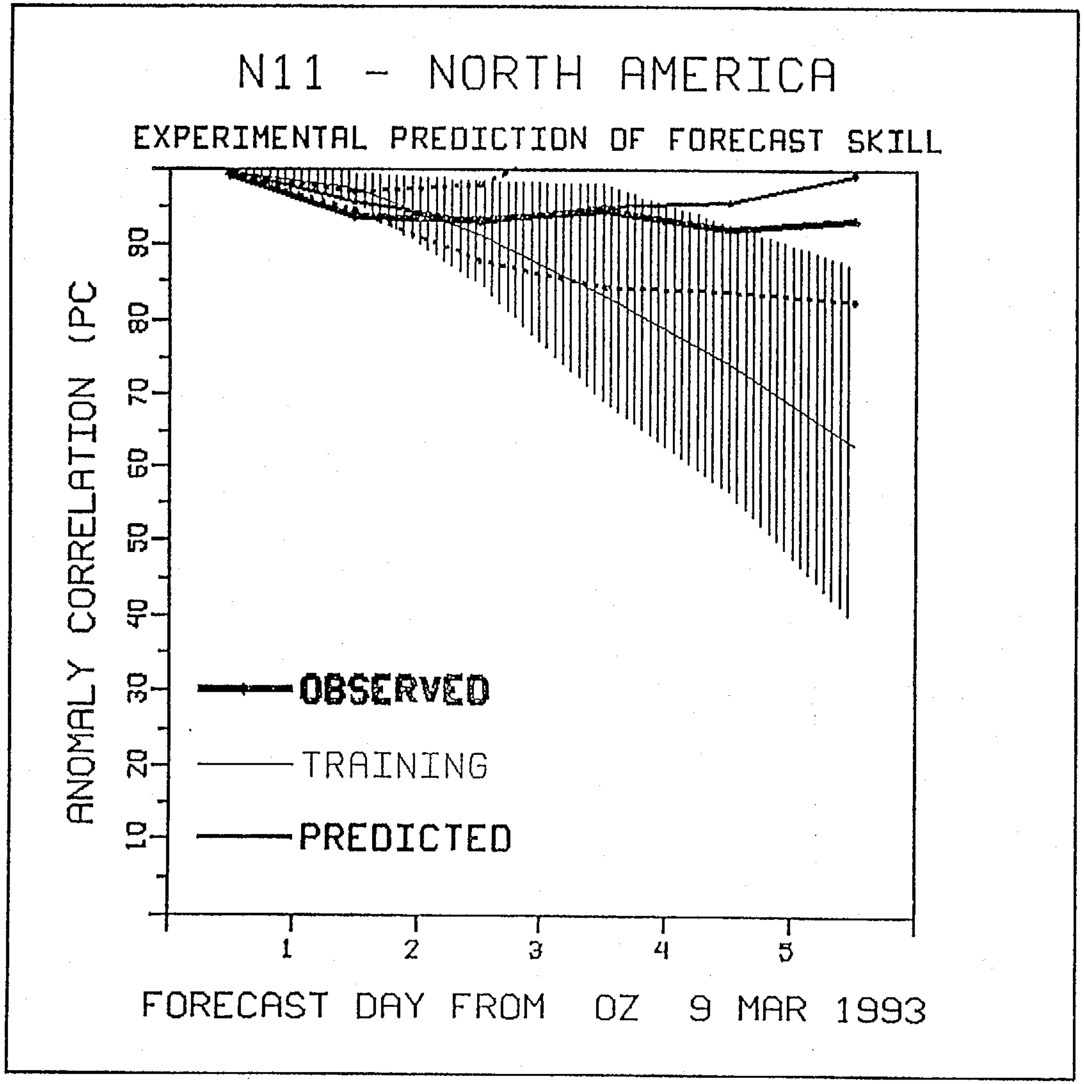

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM OZ 9 MAR 1993

Fig. 2b

Fig. 2c

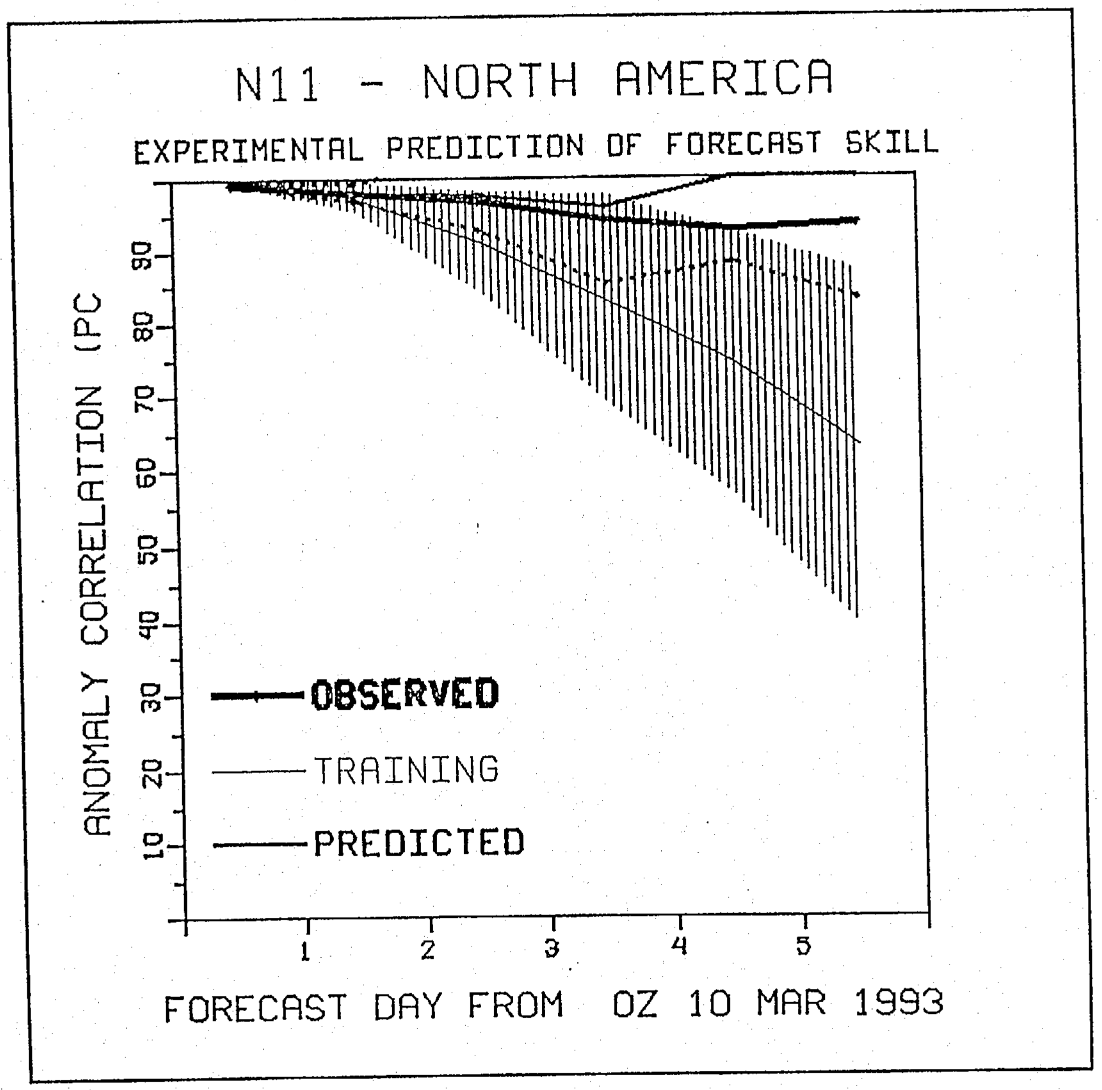

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM 0Z 10 MAR 1993

Fig. 2c

Fig. 2d

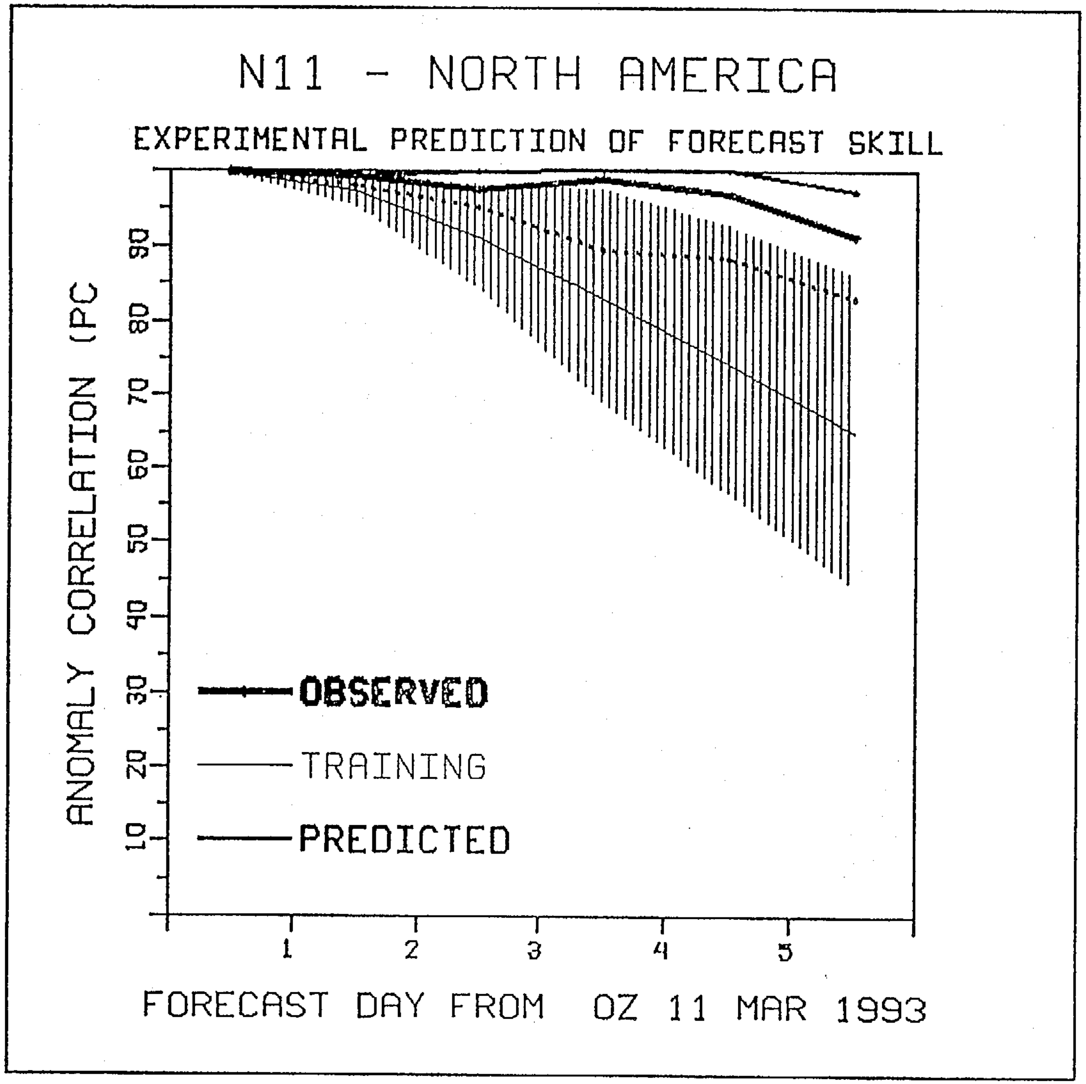

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
90
80
70
60
50
40
30
20
10
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM 0Z 11 MAR 1993

Fig. 2d

Fig. 2e

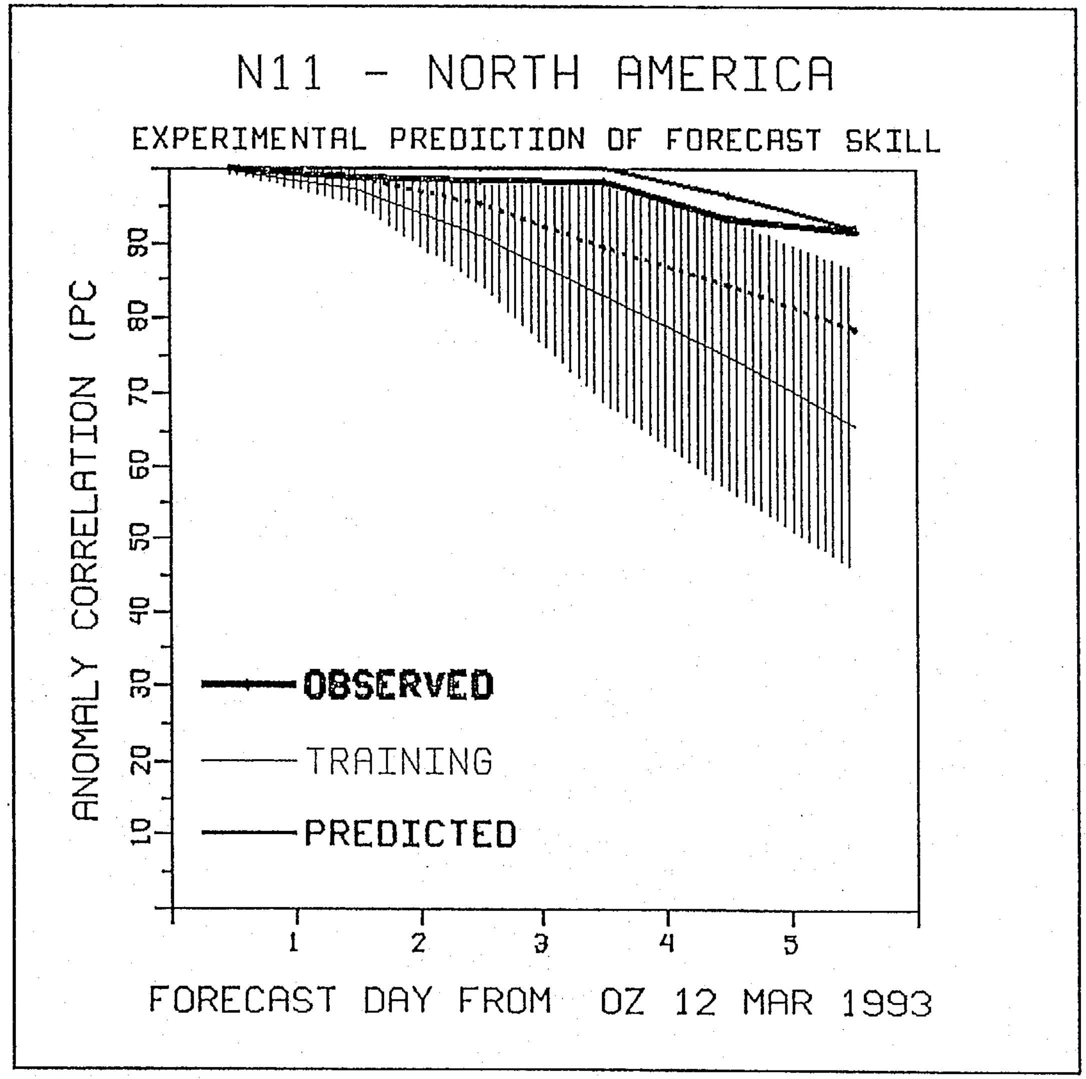

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM 0Z 12 MAR 1993

Fig. 2e

Fig. 2f

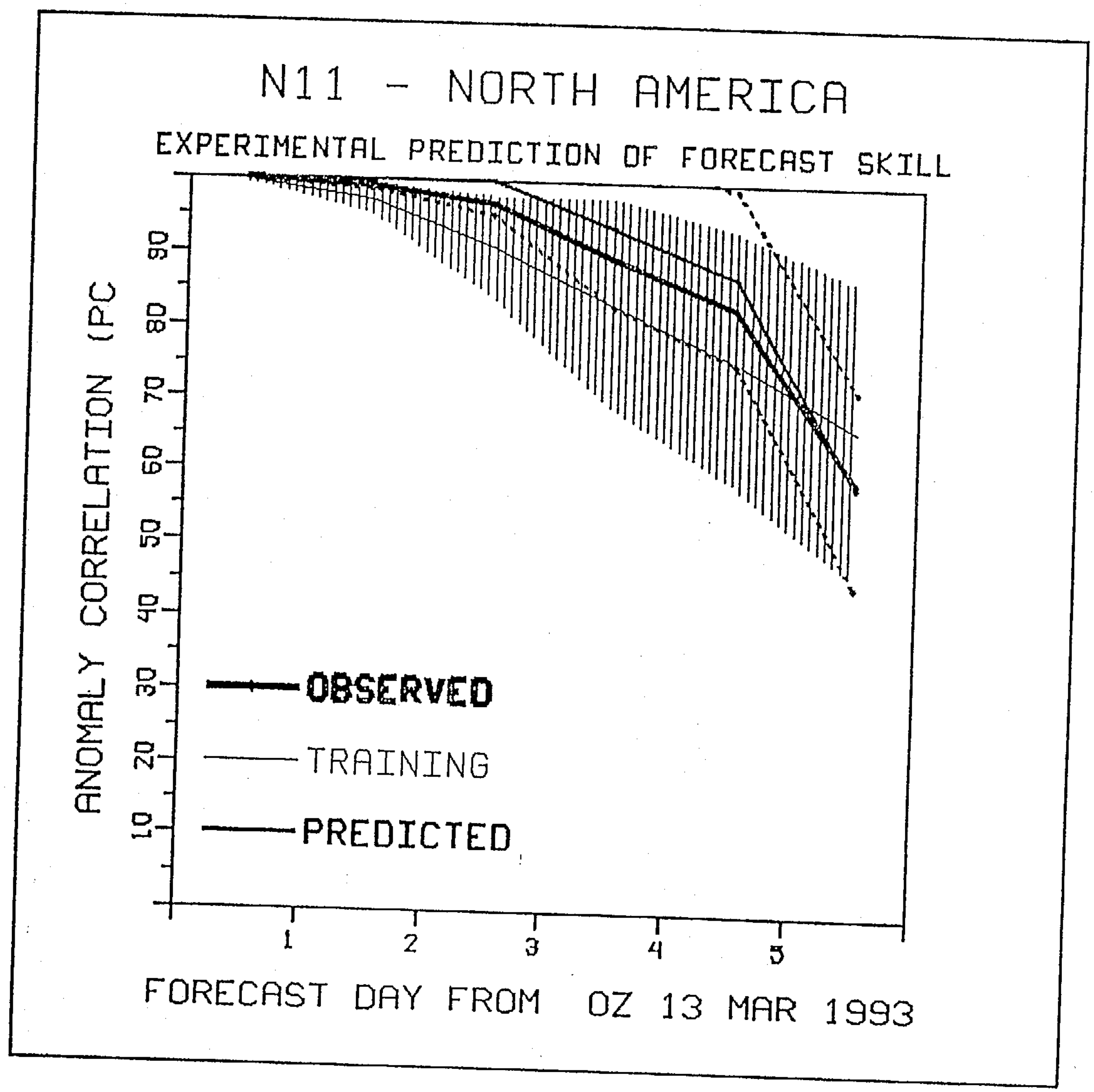

N11 - NORTH AMERICA
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM OZ 13 MAR 1993

Fig. 2f

Fig. 3a

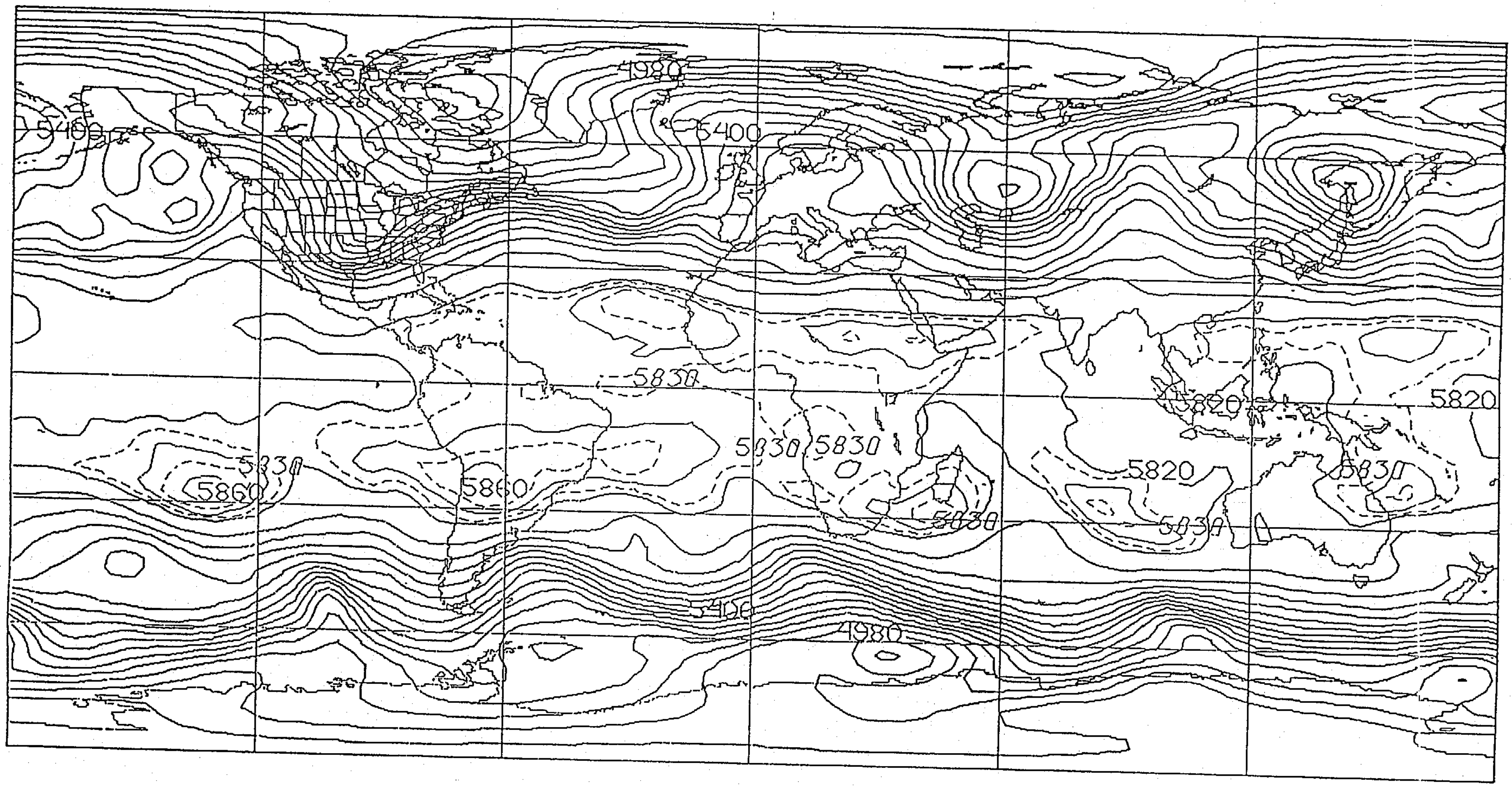
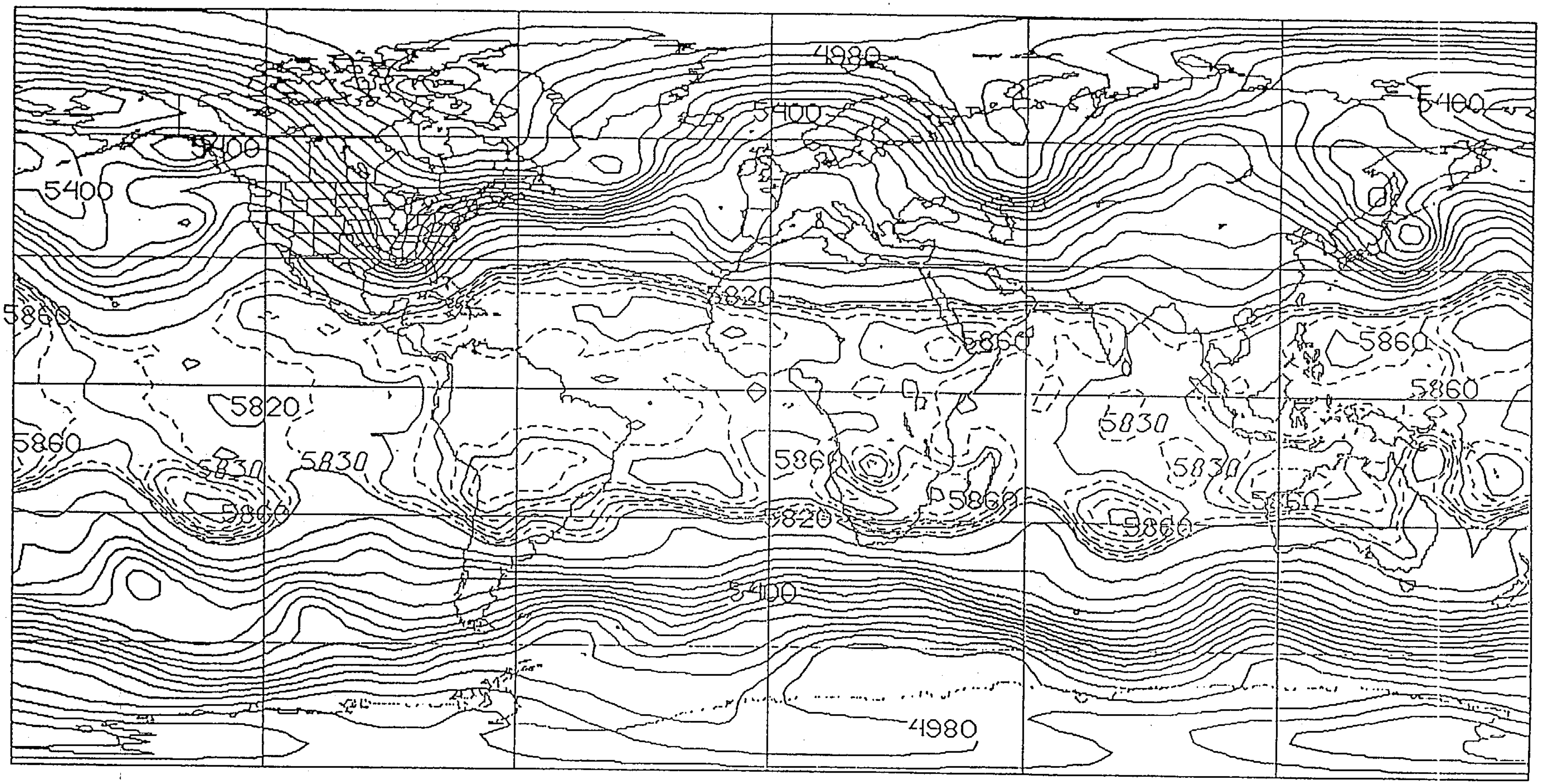

5400
1980
5400
5830
5820
5820
5830
5830
5830
5860
5860
5830
5820
5830
5830
5830
5400
4980

Fig. 3a

Fig. 3b



4980
5400
5400
5400
5400
5820
5860
5860
5860
5820
5830
5830
5860
5830
5830
5860
5860
5820
5860
5850
5400
4980

Fig. 3b

Fig.3c

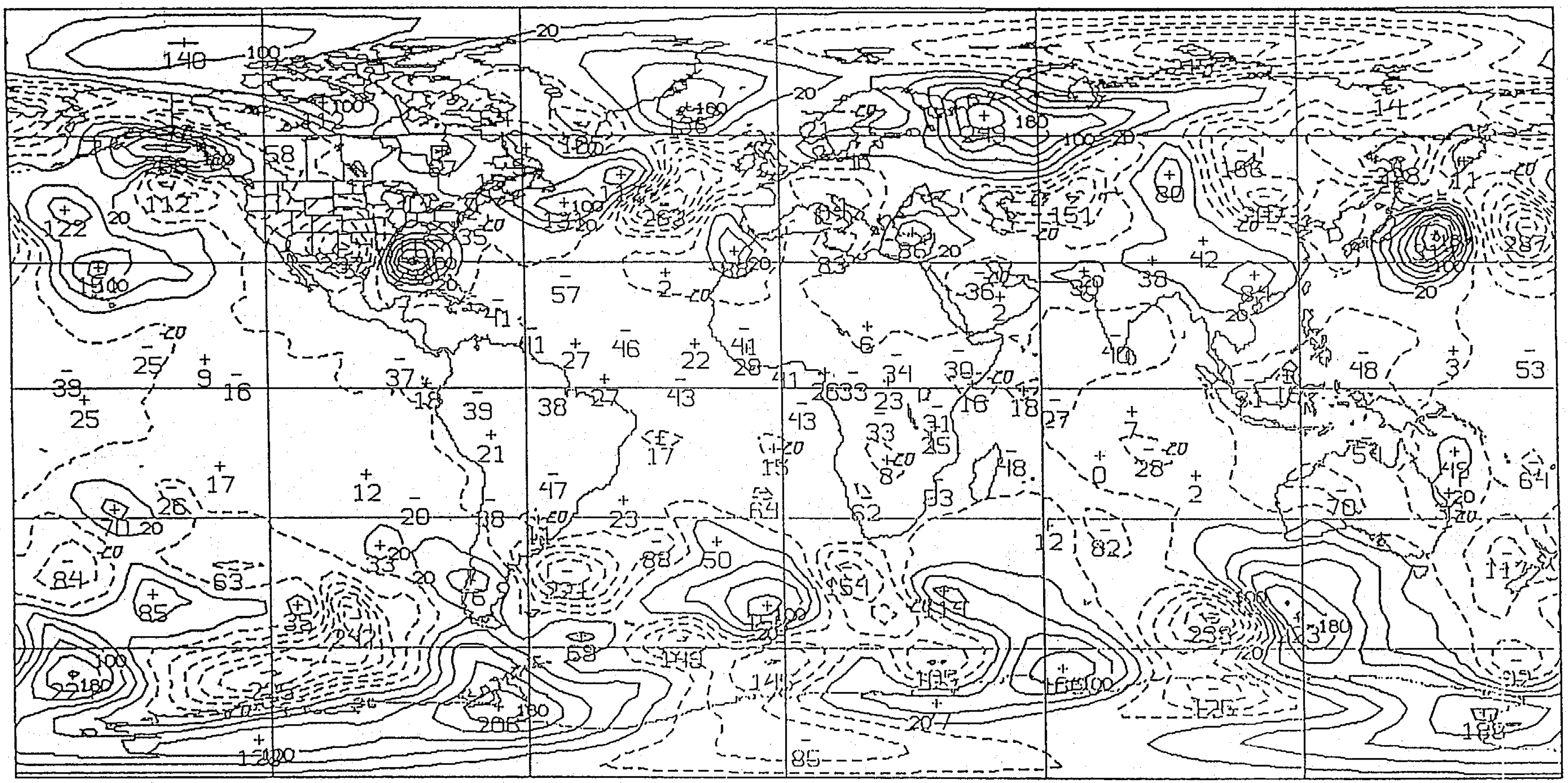
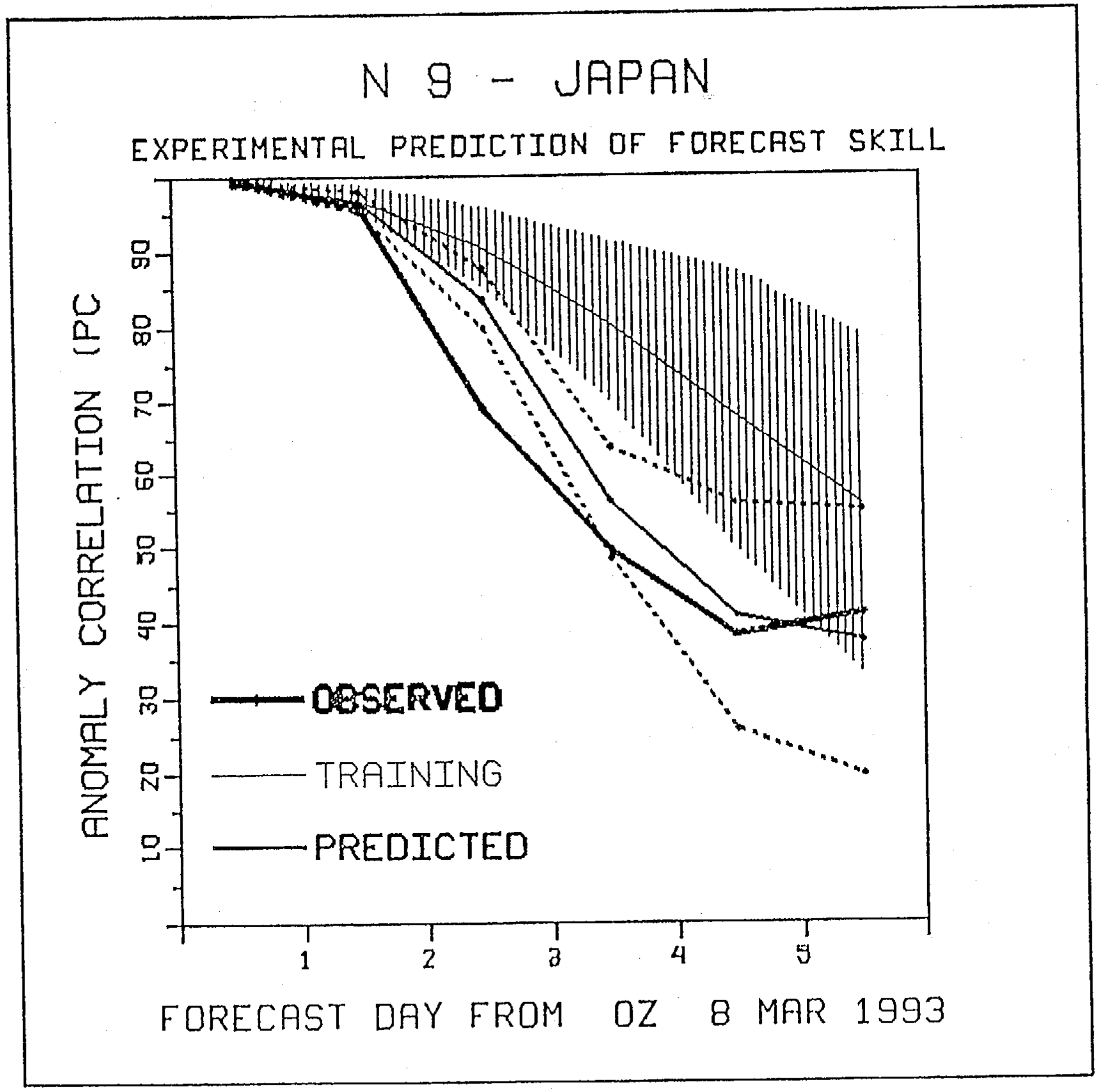


Fig. 3a

Fig. 4



N 9 - JAPAN
EXPERIMENTAL PREDICTION OF FORECAST SKILL
ANOMALY CORRELATION (PC
10
20
30
40
50
60
70
80
90
OBSERVED
TRAINING
PREDICTED
1
2
3
4
5
FORECAST DAY FROM 0Z 8 MAR 1993

Fig. 4

Fig.5a

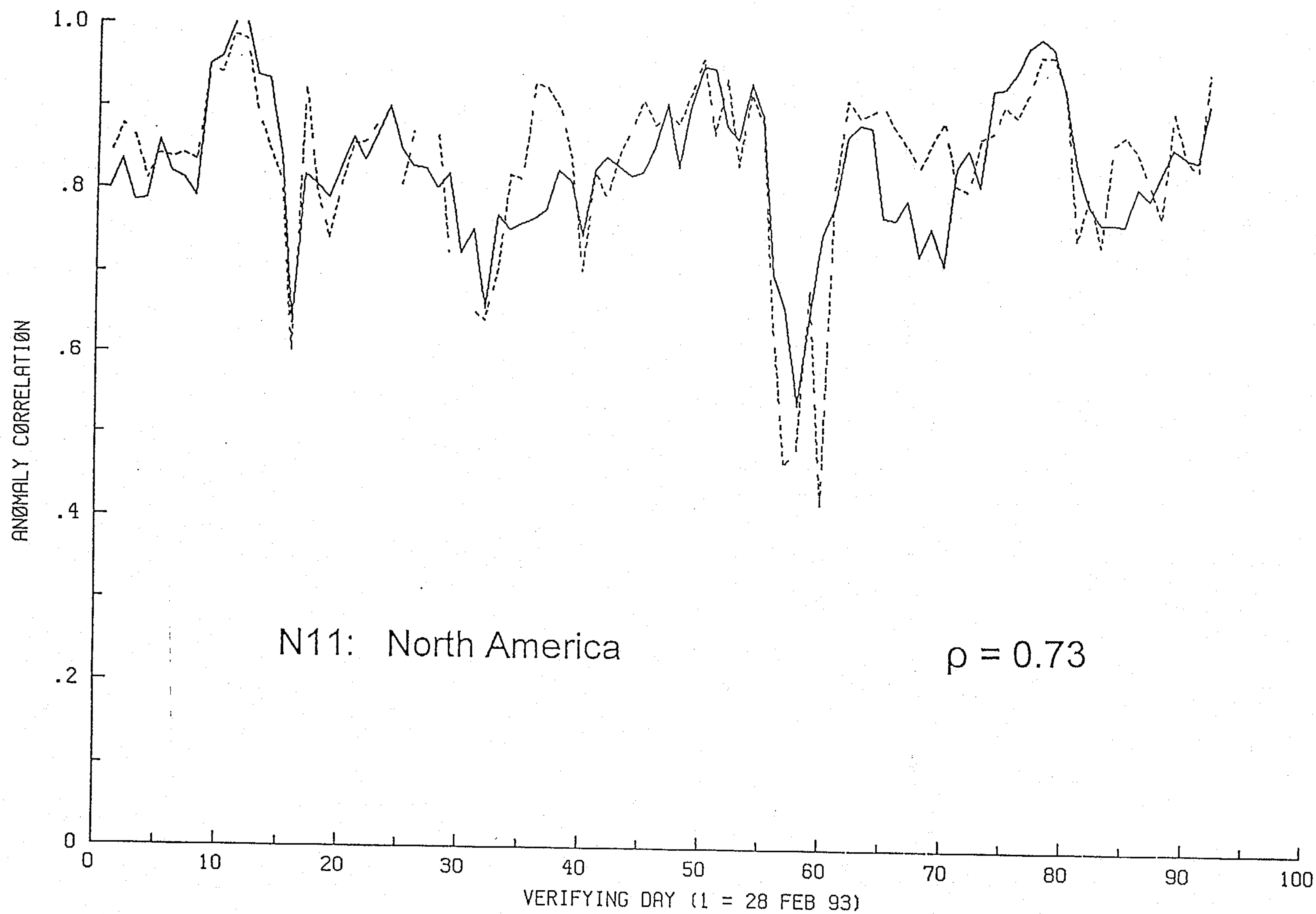
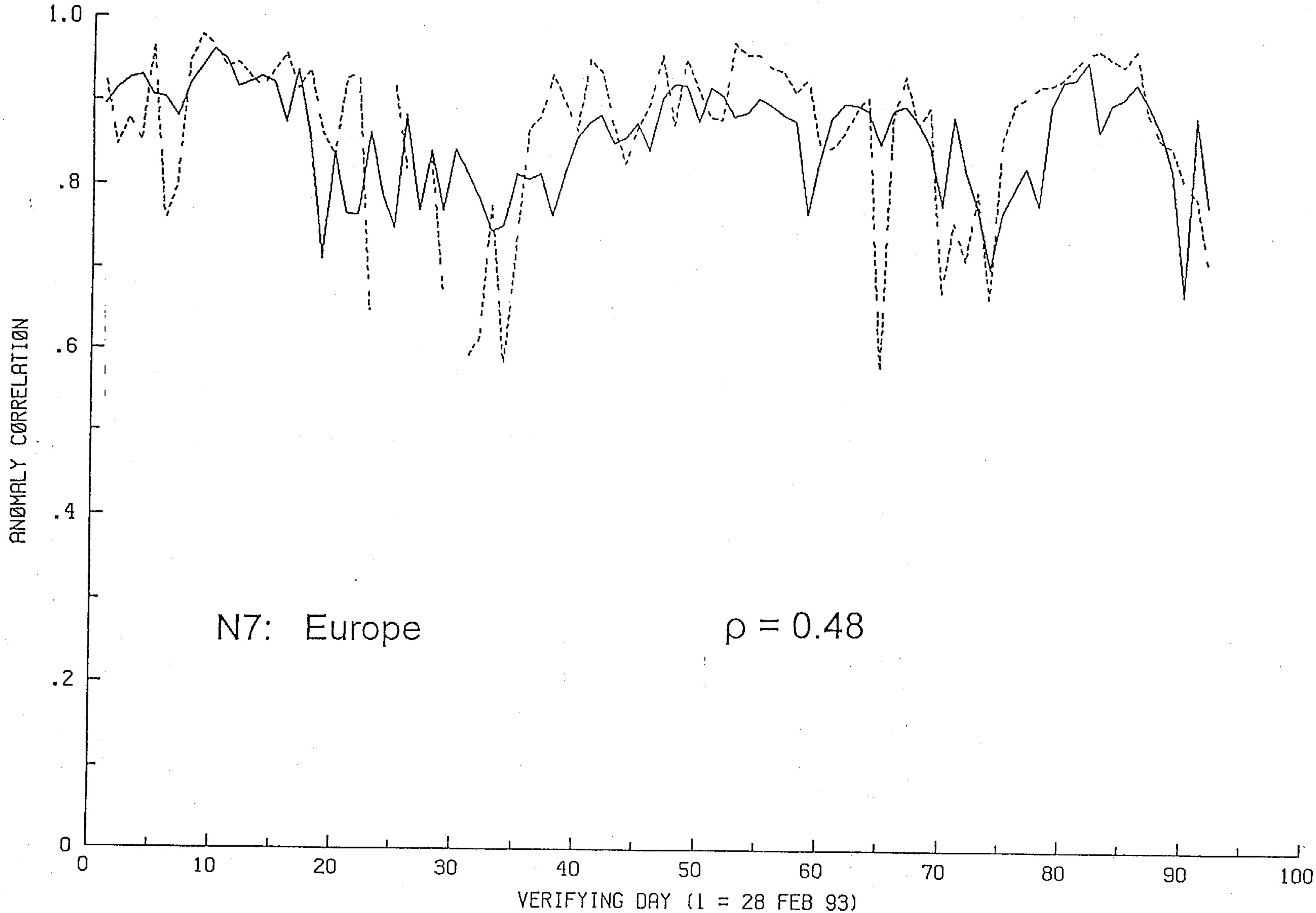

1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
N11: North America
ρ = 0.73

Fig.5a

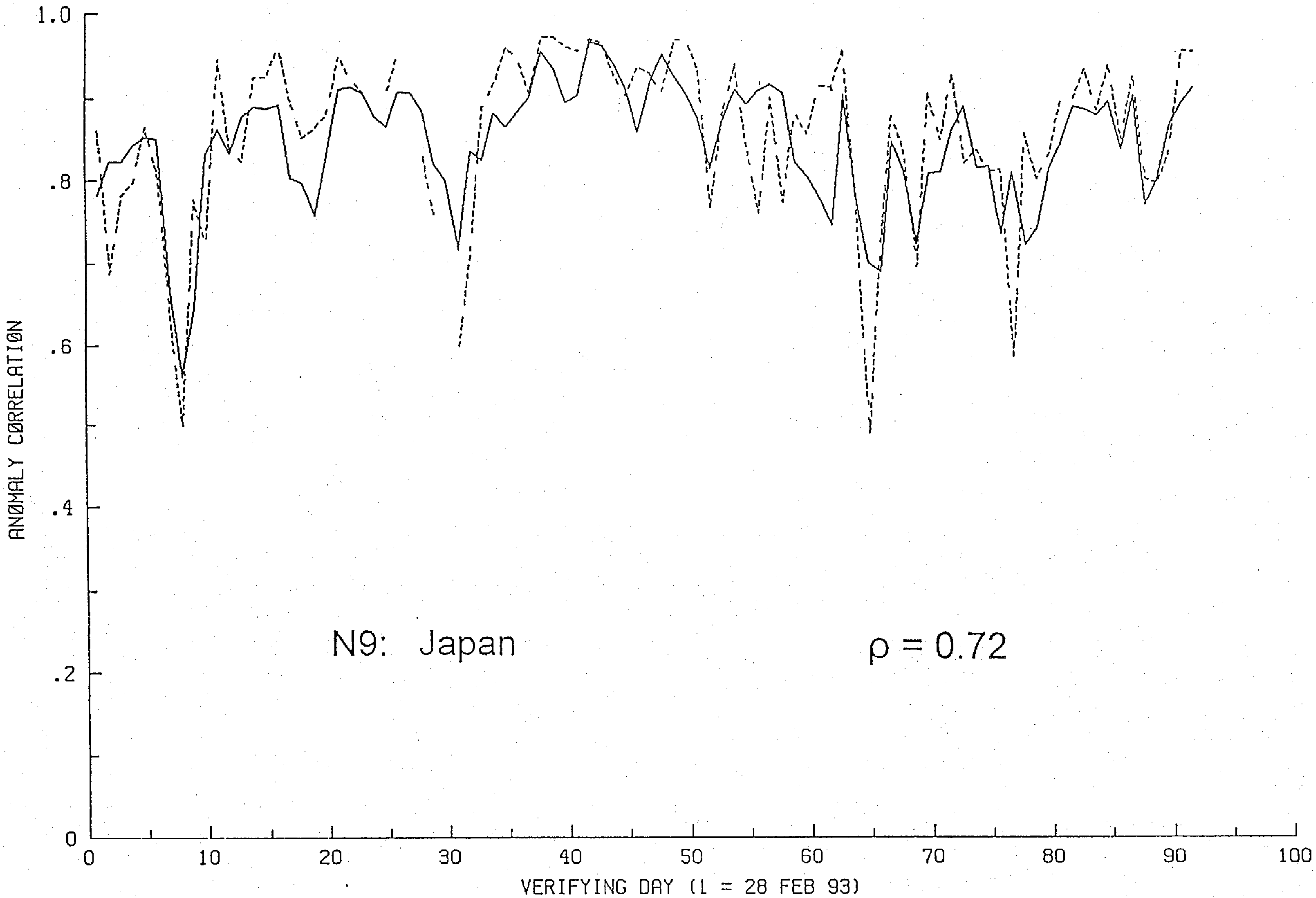
Fig. 5b



ANØMALY CØRRELATIØN
1.0
.8
.6
.4
.2
0
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
N7: Europe
ρ = 0.48

Fig. 5b

Fig. 5c



1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
N9: Japan
ρ = 0.72
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)

Fig. 5c

Fig. 6a

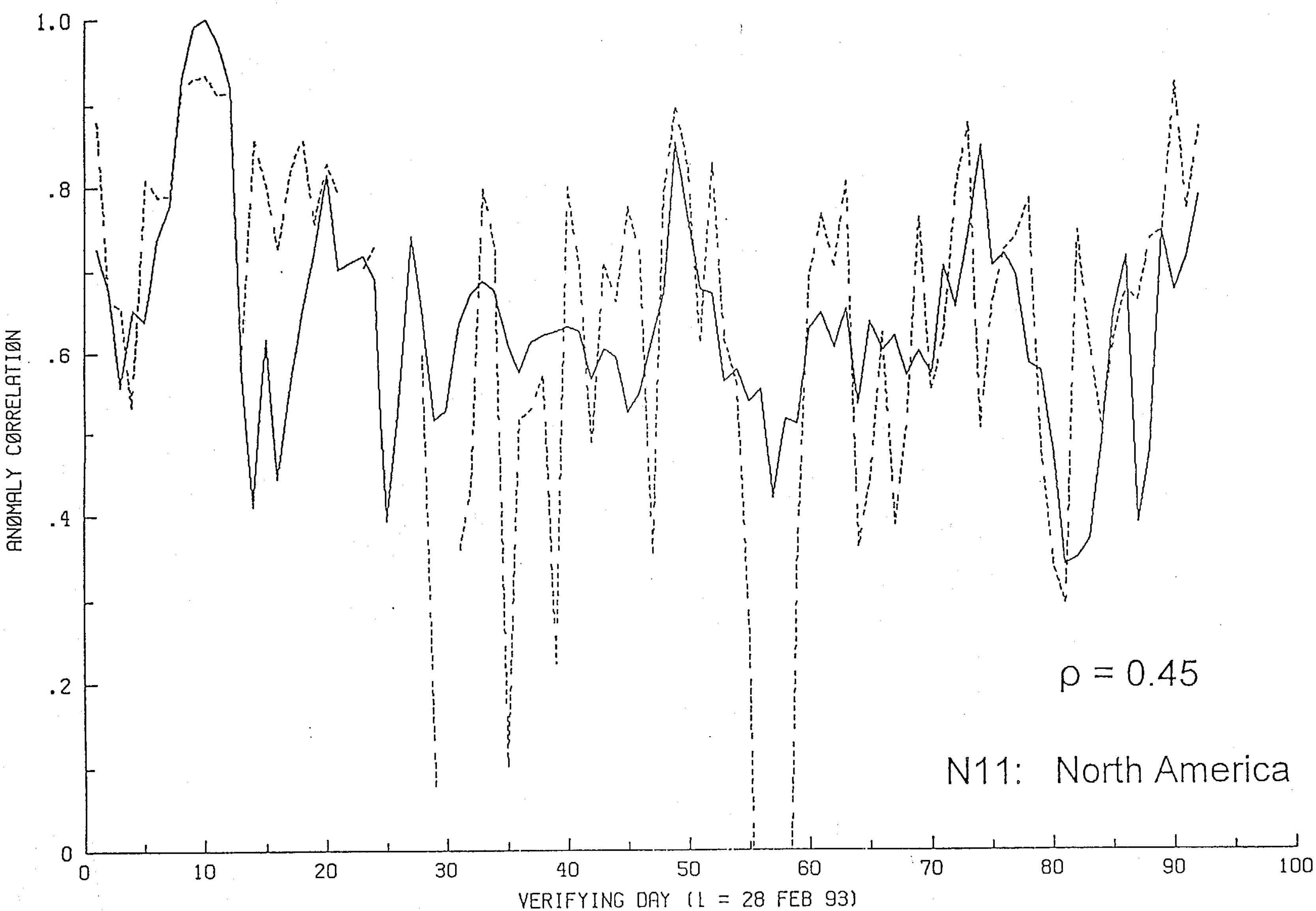
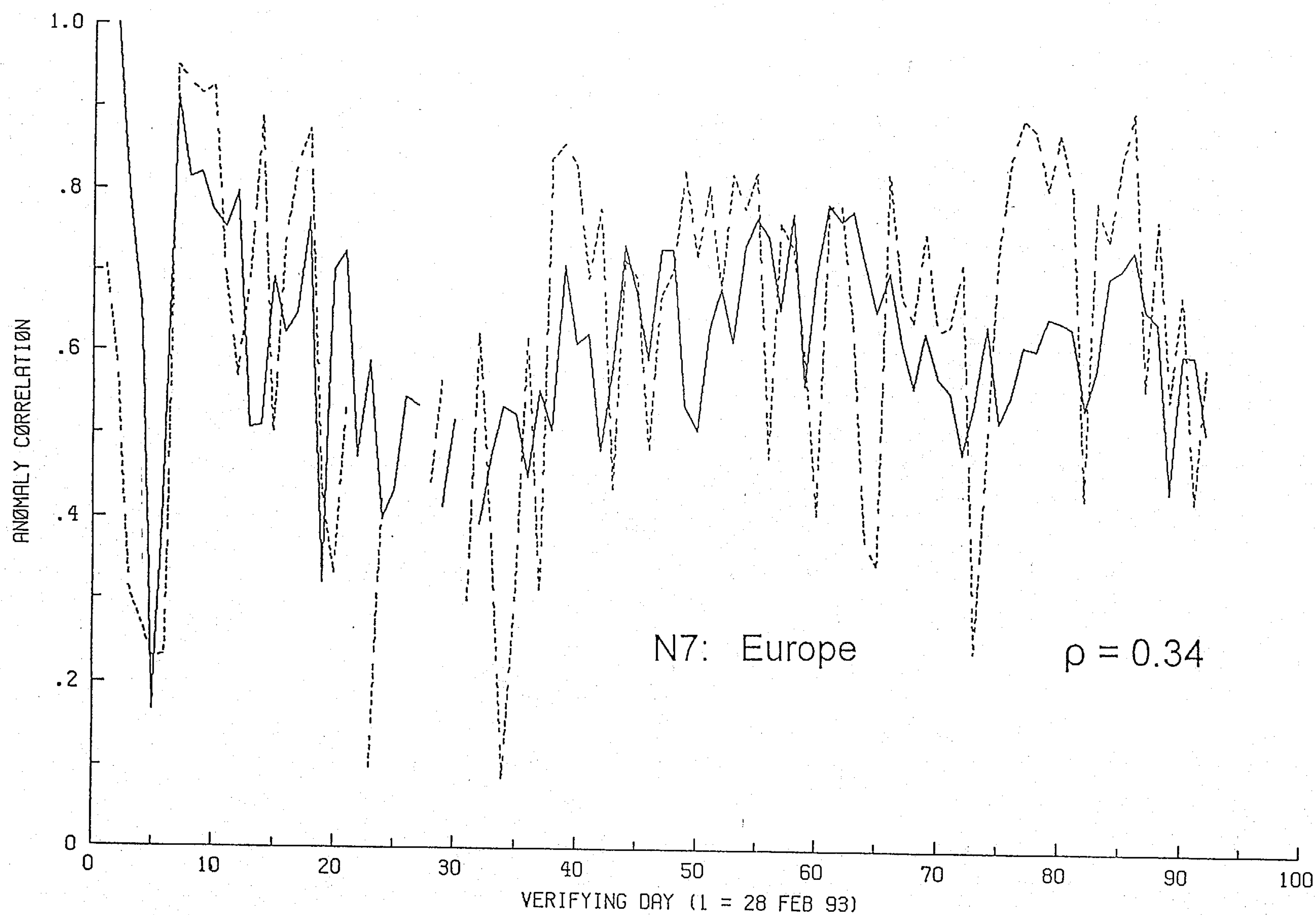

1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
ρ = 0.45
N11: North America

Fig. 6a

Fig. 6b



1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
N7: Europe
ρ = 0.34

Fig. 6b

Fig. 6c.

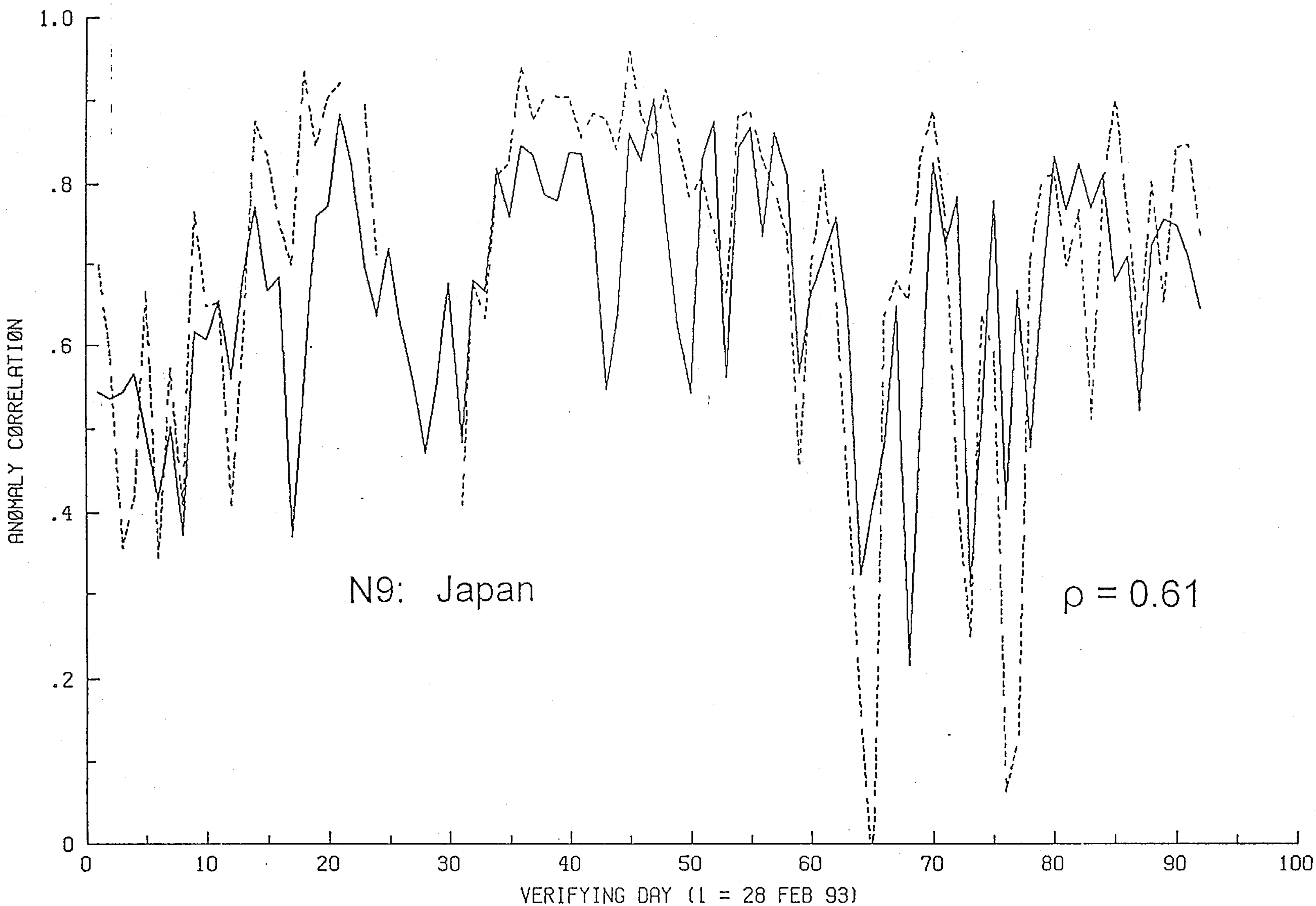
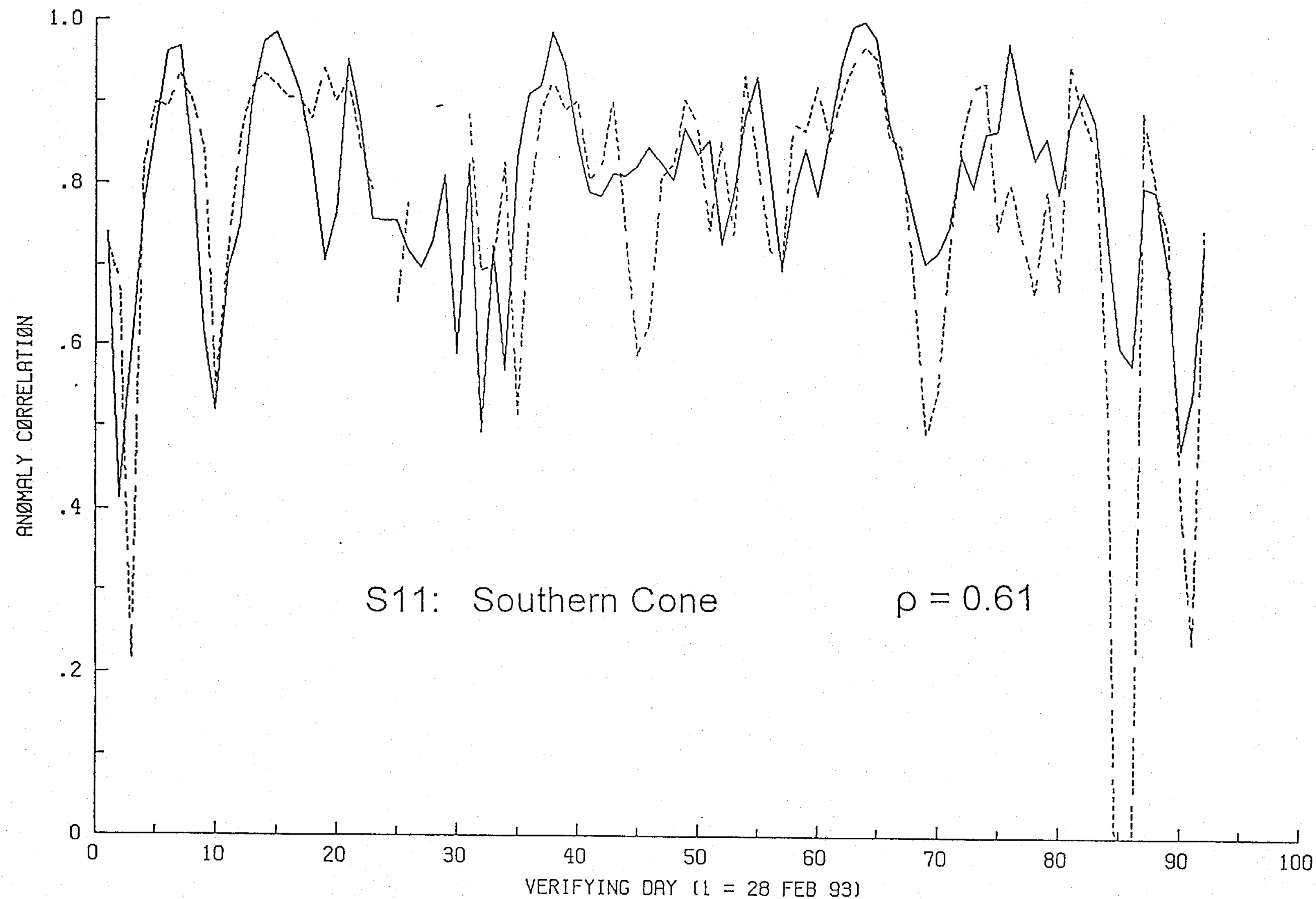

1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
N9: Japan
ρ = 0.61

Fig. 6c'

Fig. 7a



1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
S11: Southern Cone
ρ = 0.61

Fig. 7a

Fig. 7b

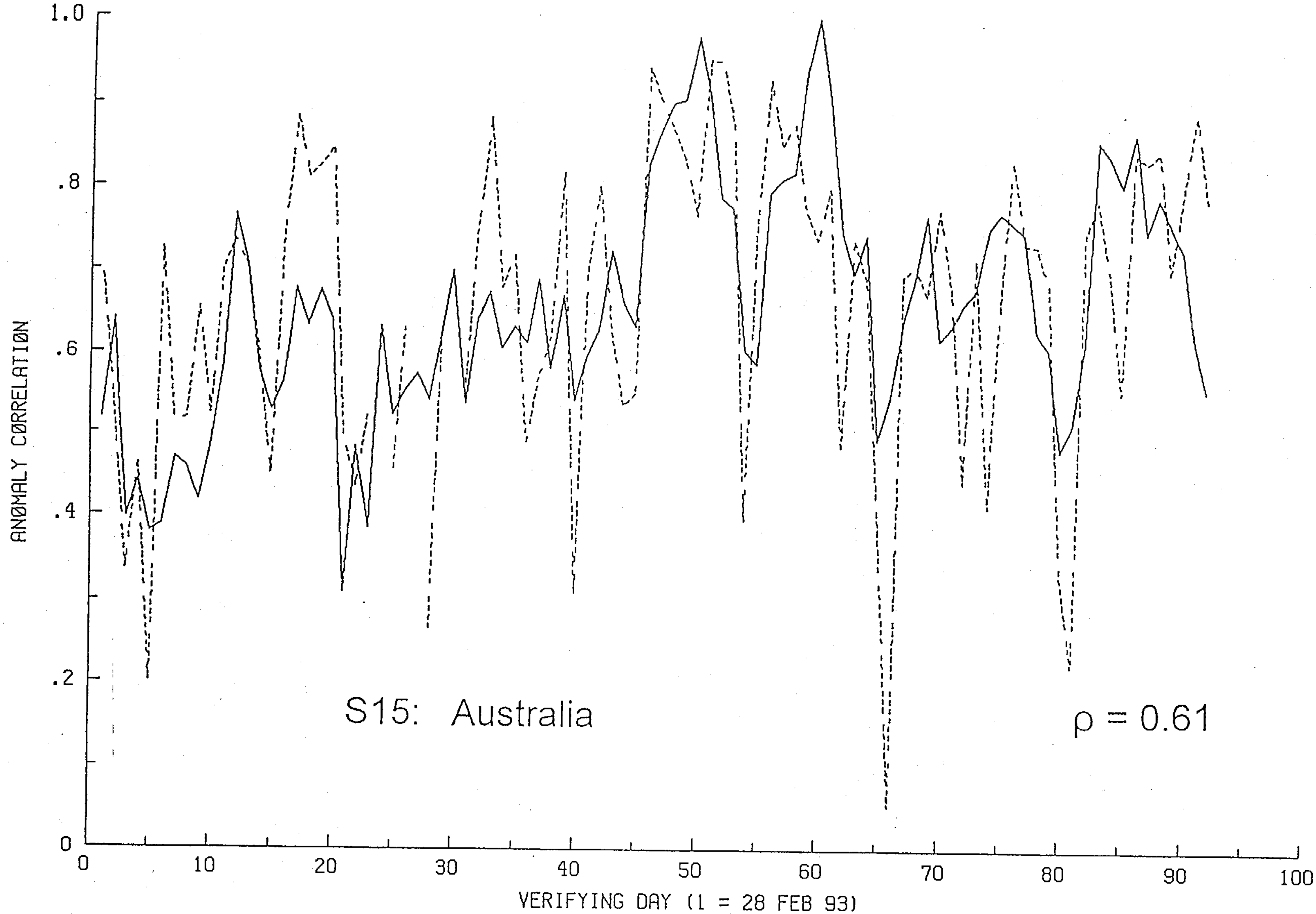
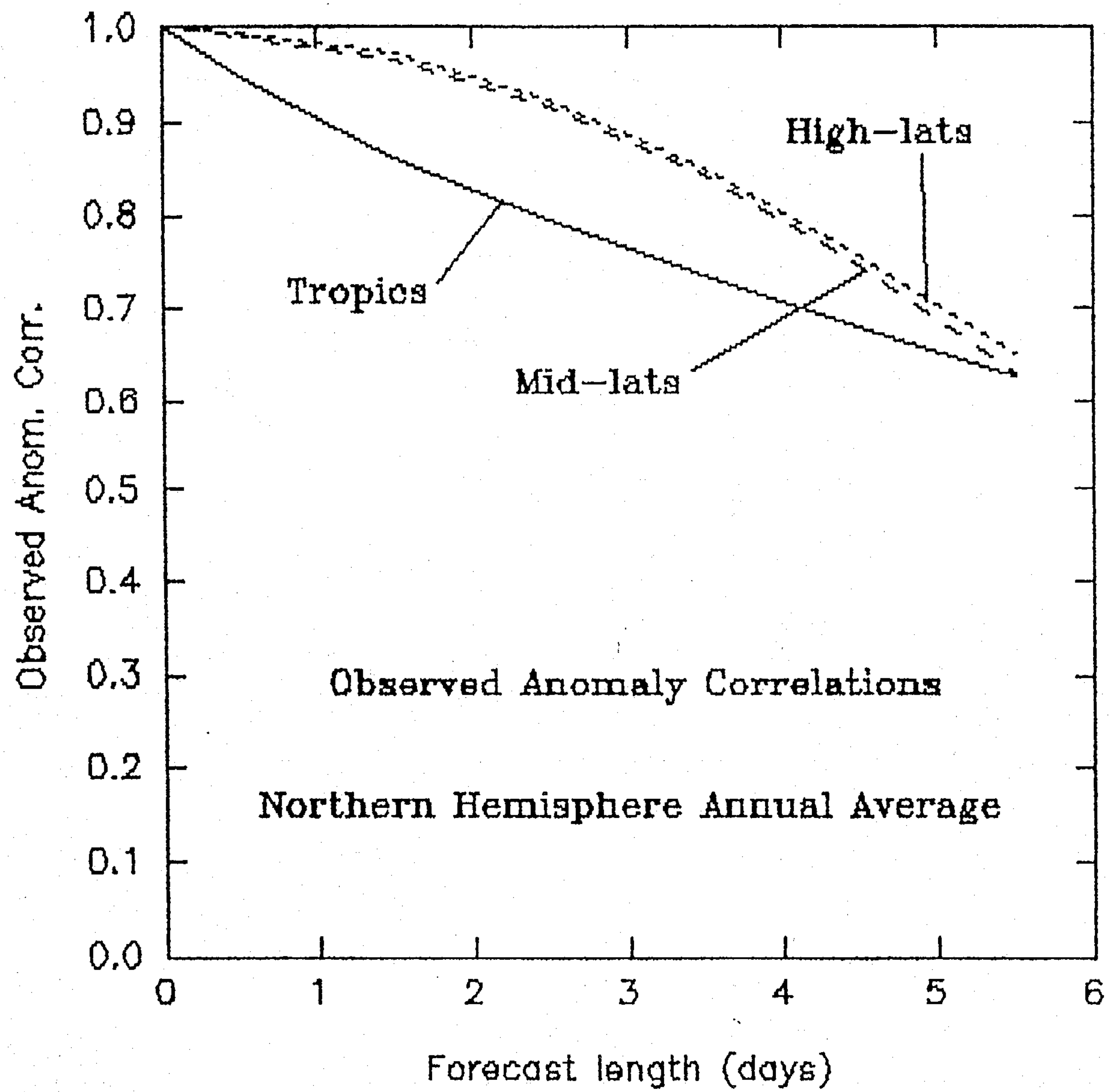

1.0
.8
.6
.4
.2
0
ANØMALY CØRRELATIØN
0
10
20
30
40
50
60
70
80
90
100
VERIFYING DAY (1 = 28 FEB 93)
S15: Australia
ρ = 0.61

Fig. 7b

Fig. 8



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
Observed Anom. Corr.
Forecast length (days)
High-lats
Tropics
Mid-lats
Observed Anomaly Correlations
Northern Hemisphere Annual Average

Fig. 8

Fig. 9

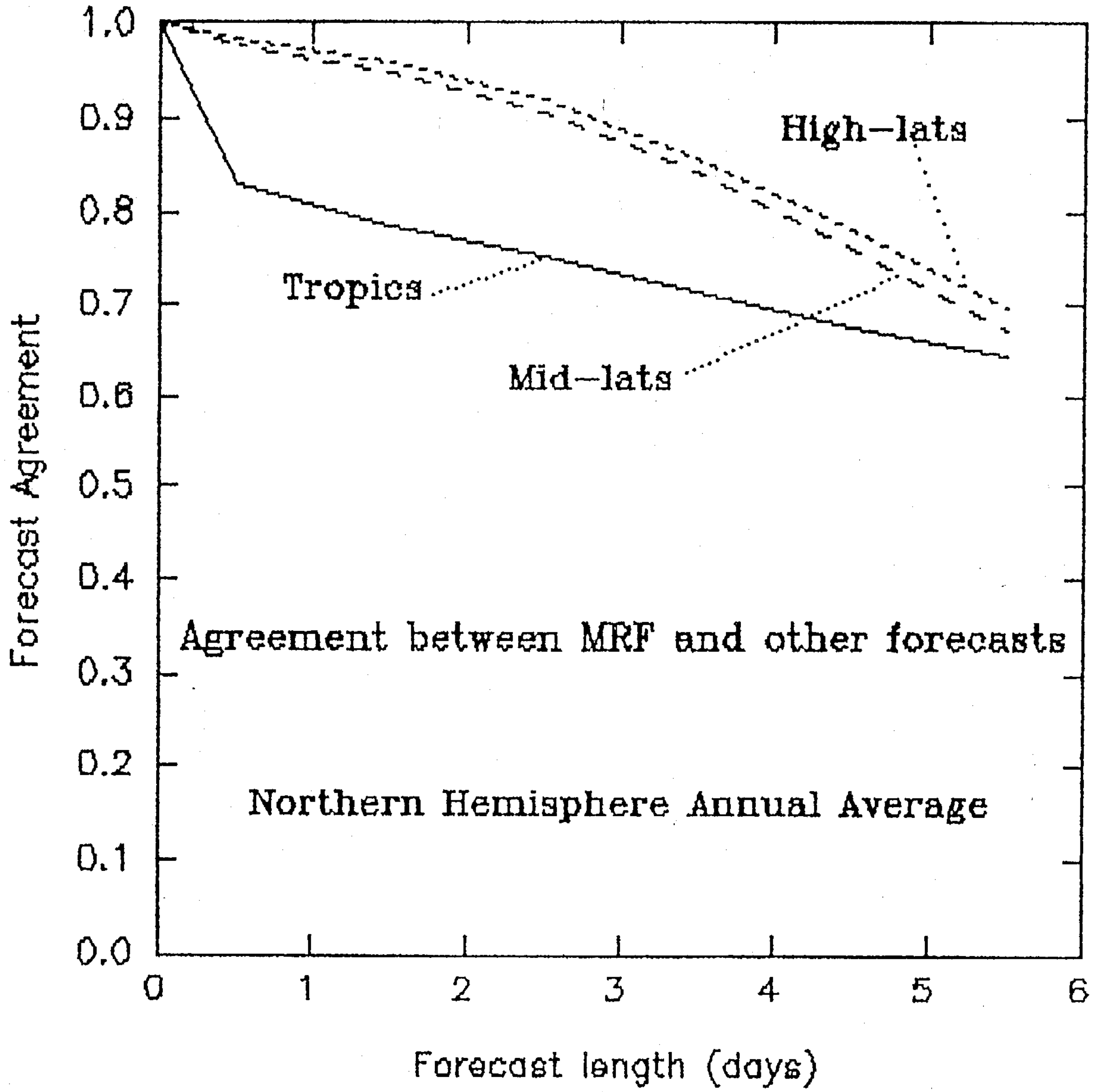

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
Forecast Agreement
Forecast length (days)
High-lats
Tropics
Mid-lats
Agreement between MRF and other forecasts
Northern Hemisphere Annual Average

Fig. 9

Fig. 10a

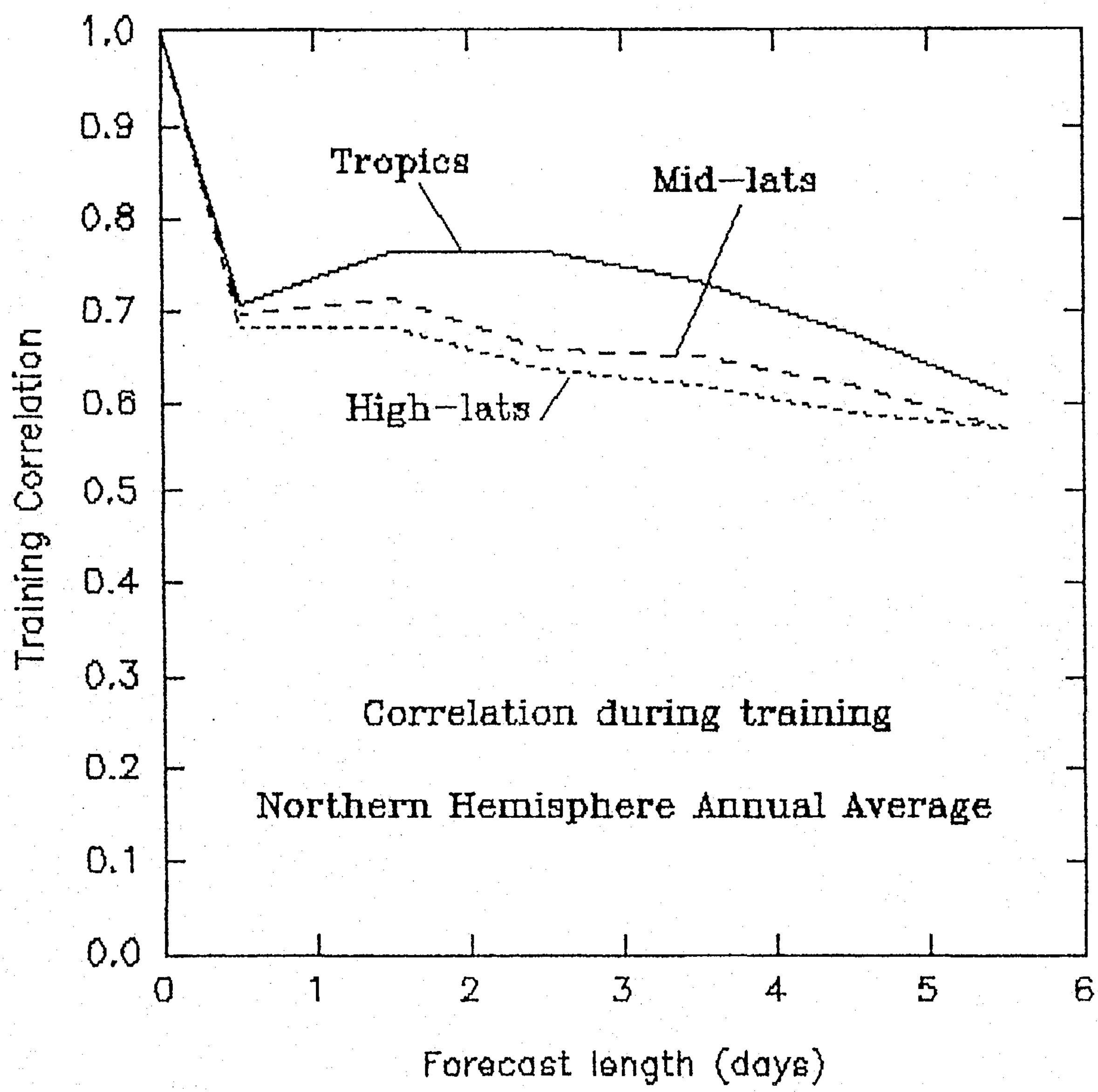


Fig. 10a

Fig. 10b

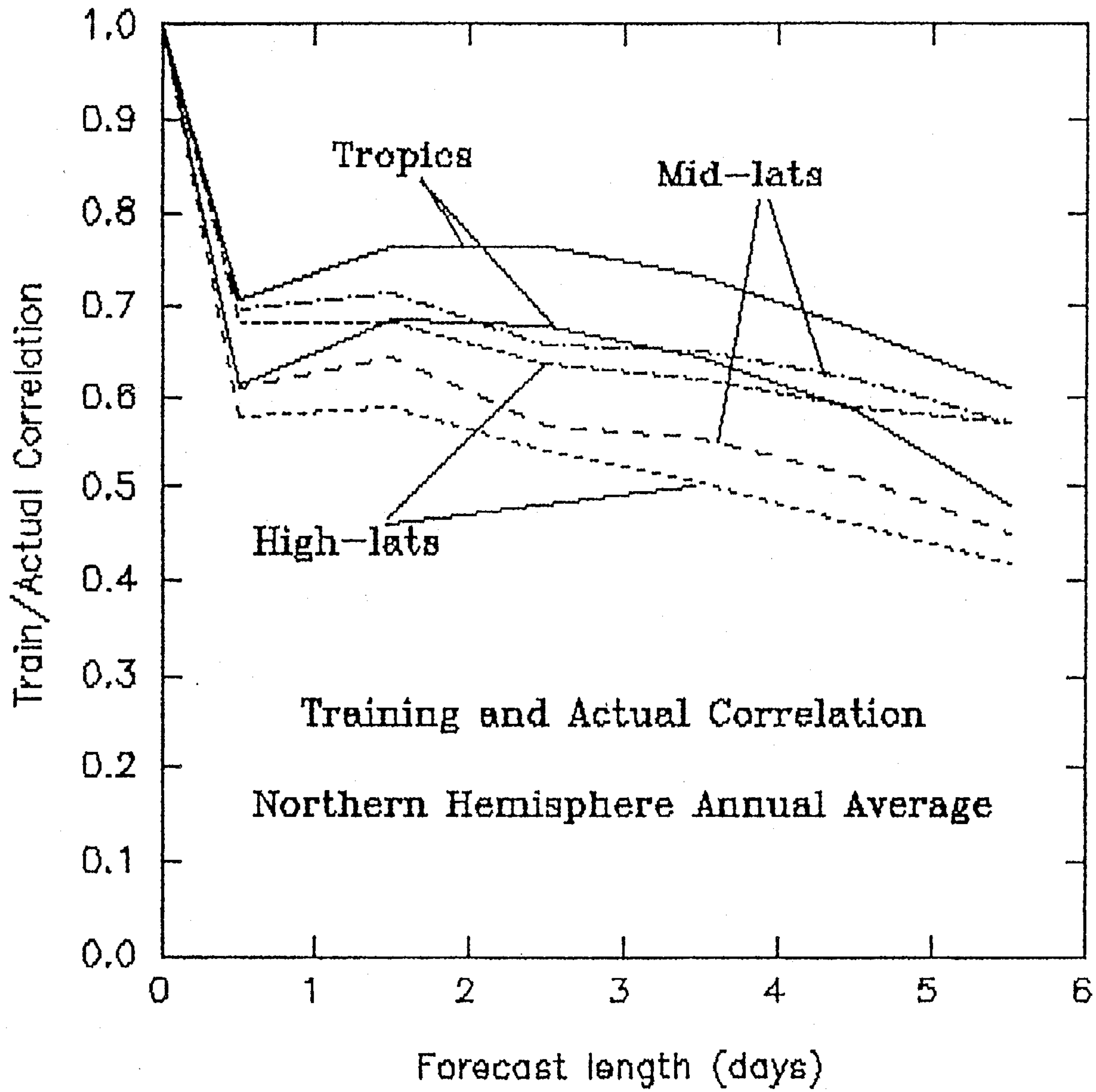


Fig. 10b

Fig. 10c

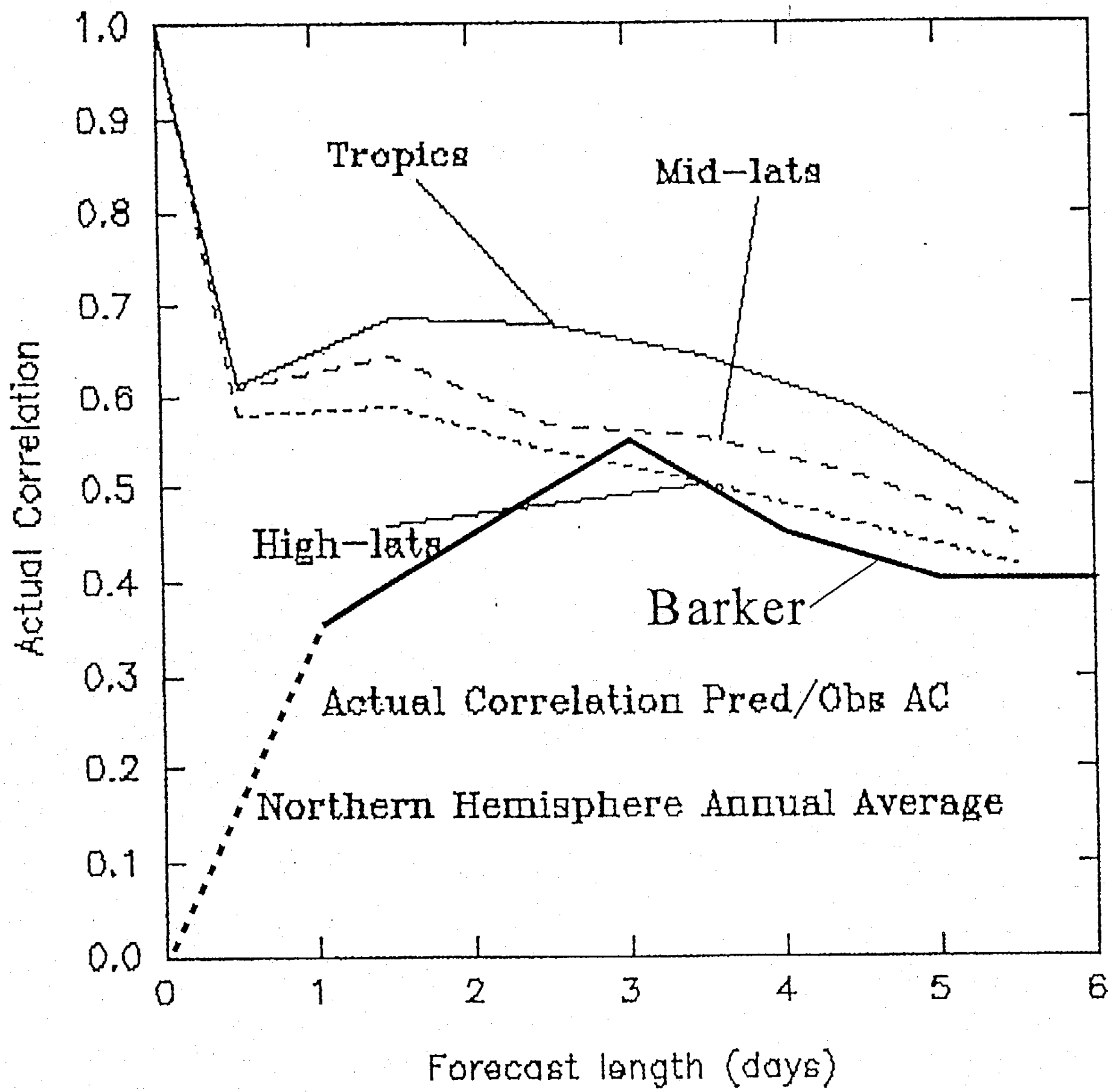


Fig. 10c

Fig. 11

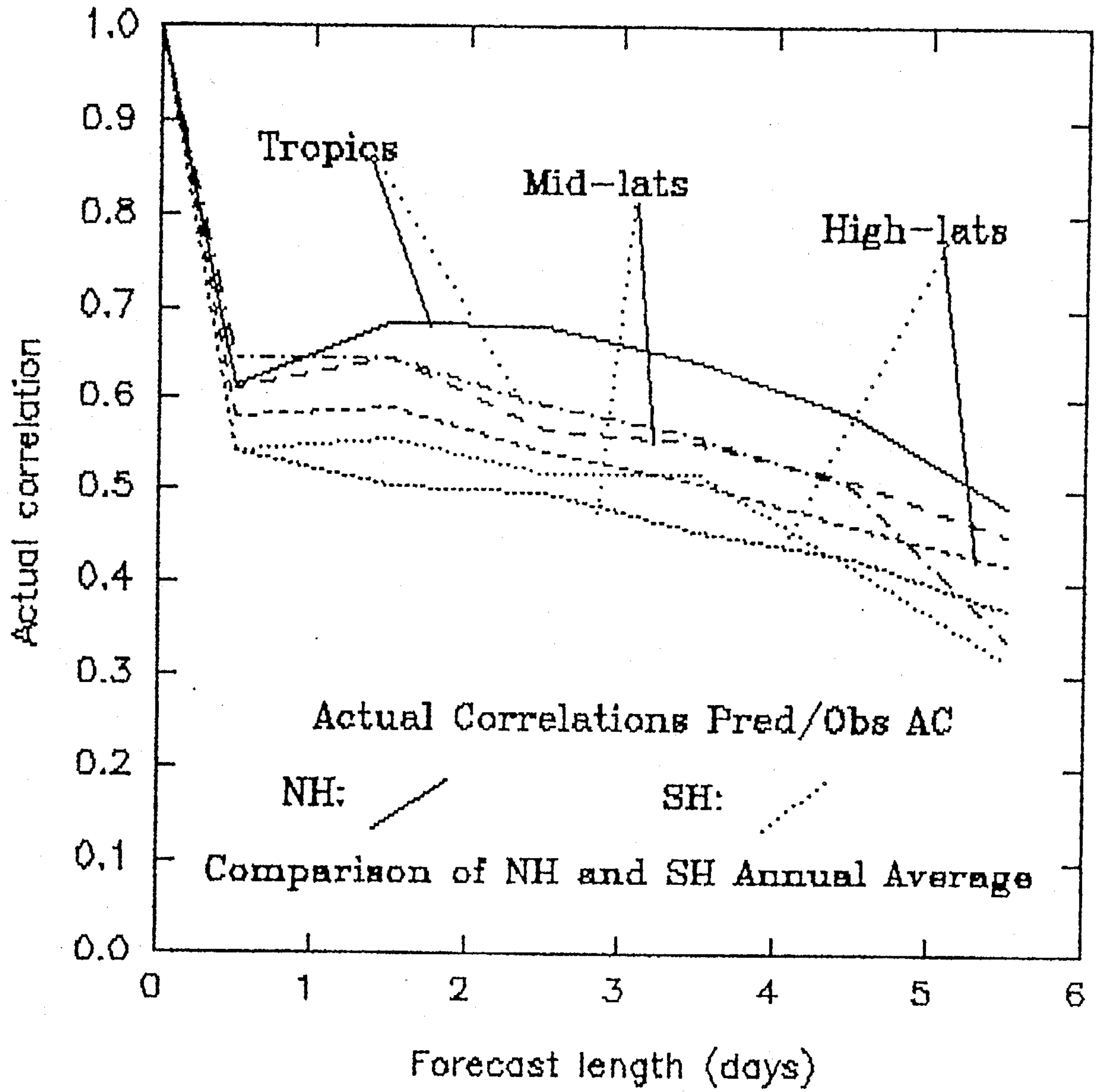

Tropics
Mid-lats
High-lats
Actual Correlations Pred/Obs AC
NH:
SH:
Comparison of NH and SH Annual Average
Actual correlation
Forecast length (days)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6

Fig. 11

Fig. 12a

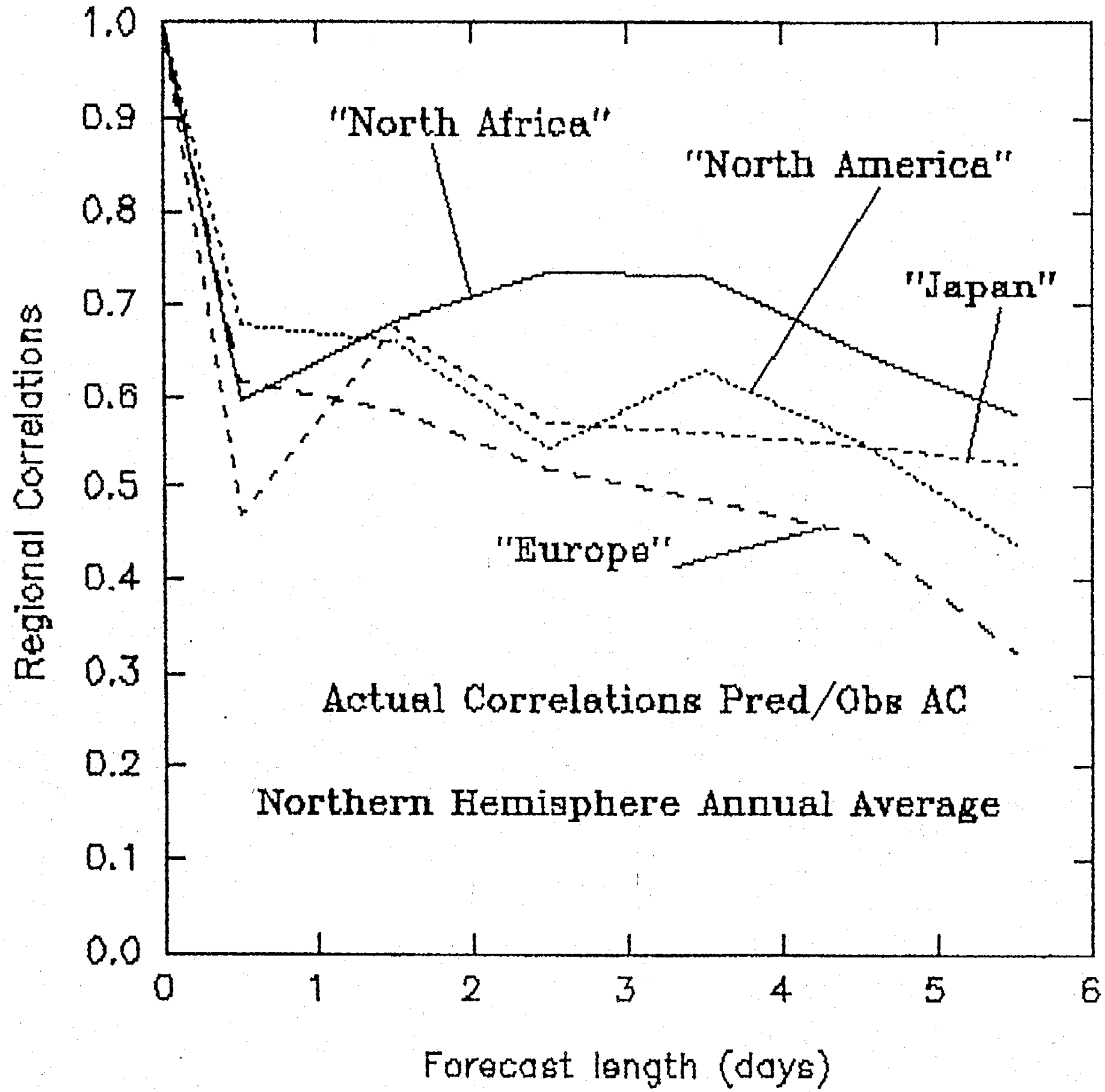


Fig. 12 a

Fig. 12b'

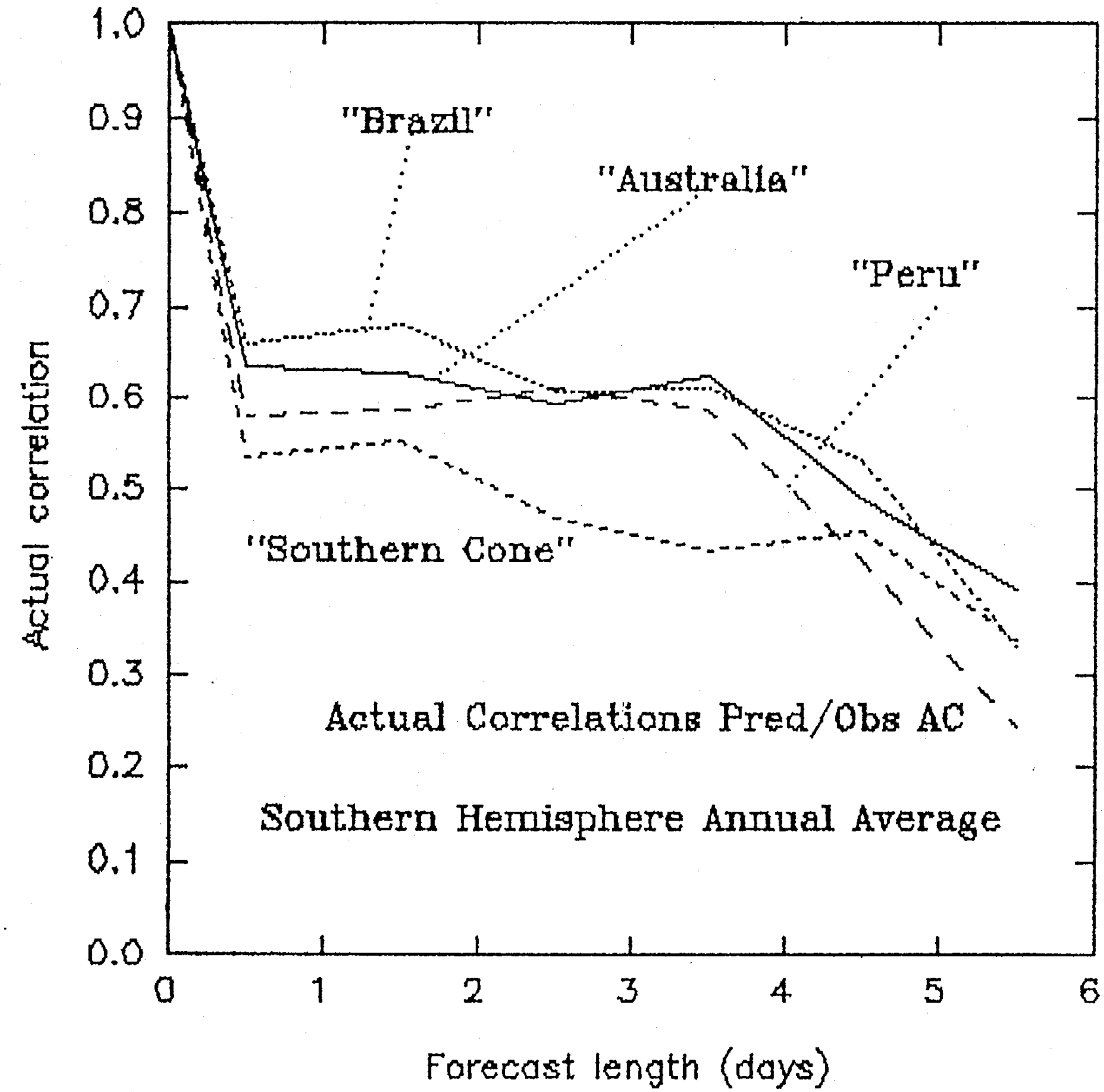
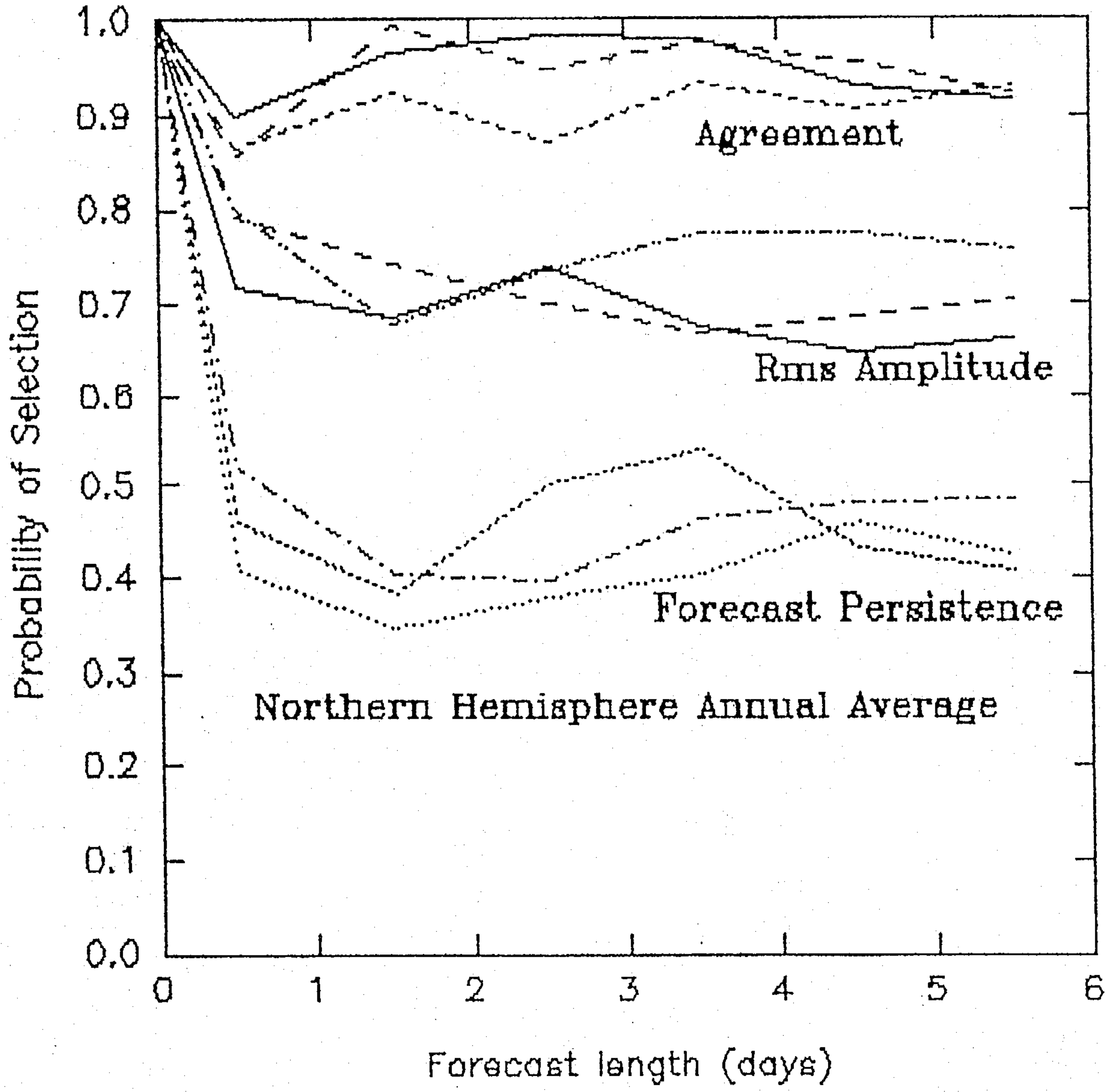


Fig. 12b'

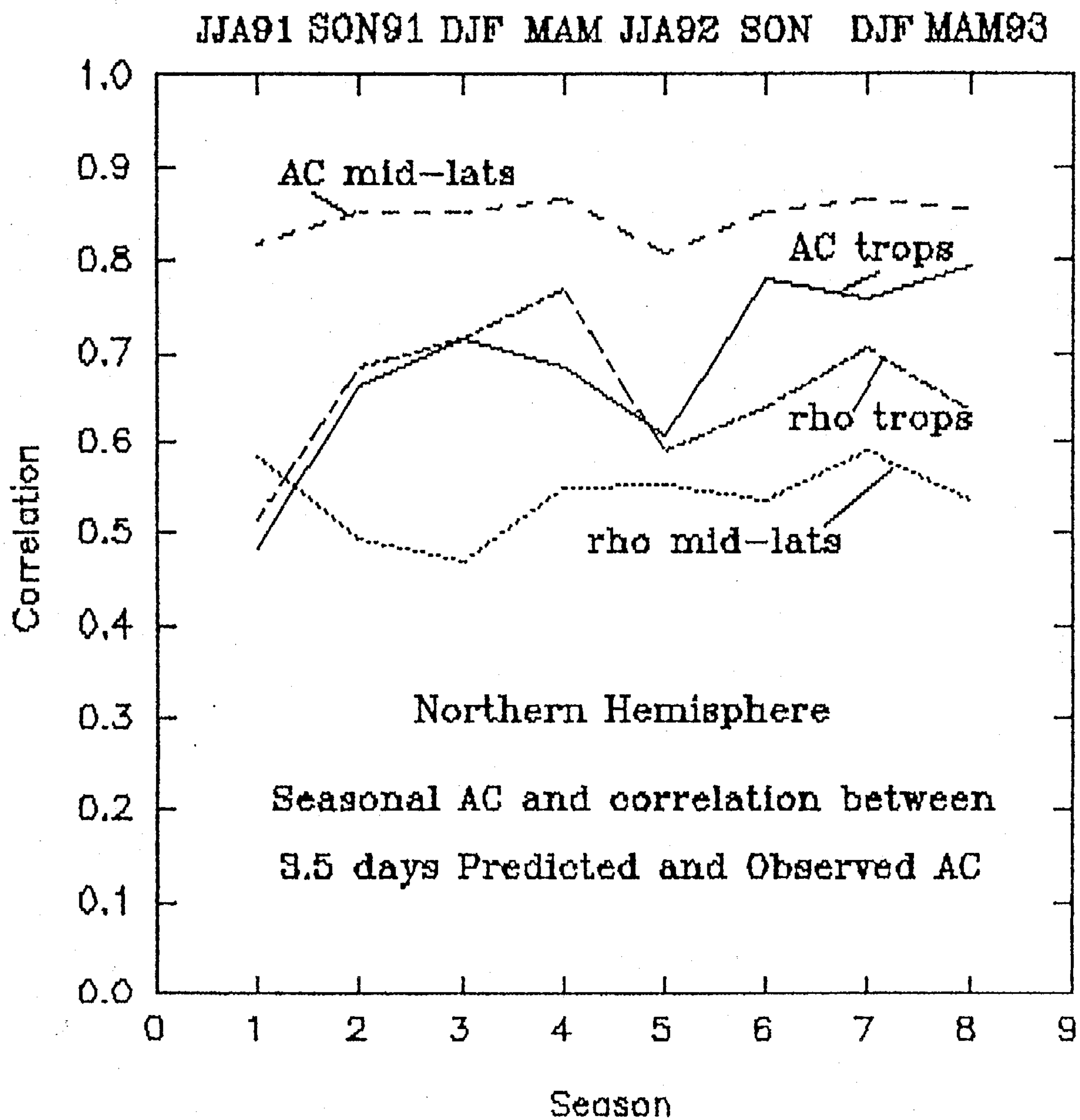
Fig. 13



Probability of Selection
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
Forecast length (days)
Agreement
Rms Amplitude
Forecast Persistence
Northern Hemisphere Annual Average

Fig. 13

Fig. 14



JJA91 SON91 DJF MAM JJA92 SON DJF MAM93
AC mid-lats
AC trops
rho trops
rho mid-lats
Northern Hemisphere
Seasonal AC and correlation between
3.5 days Predicted and Observed AC
Correlation
Season
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
7
8
9

Fig. 14

Fig.15a

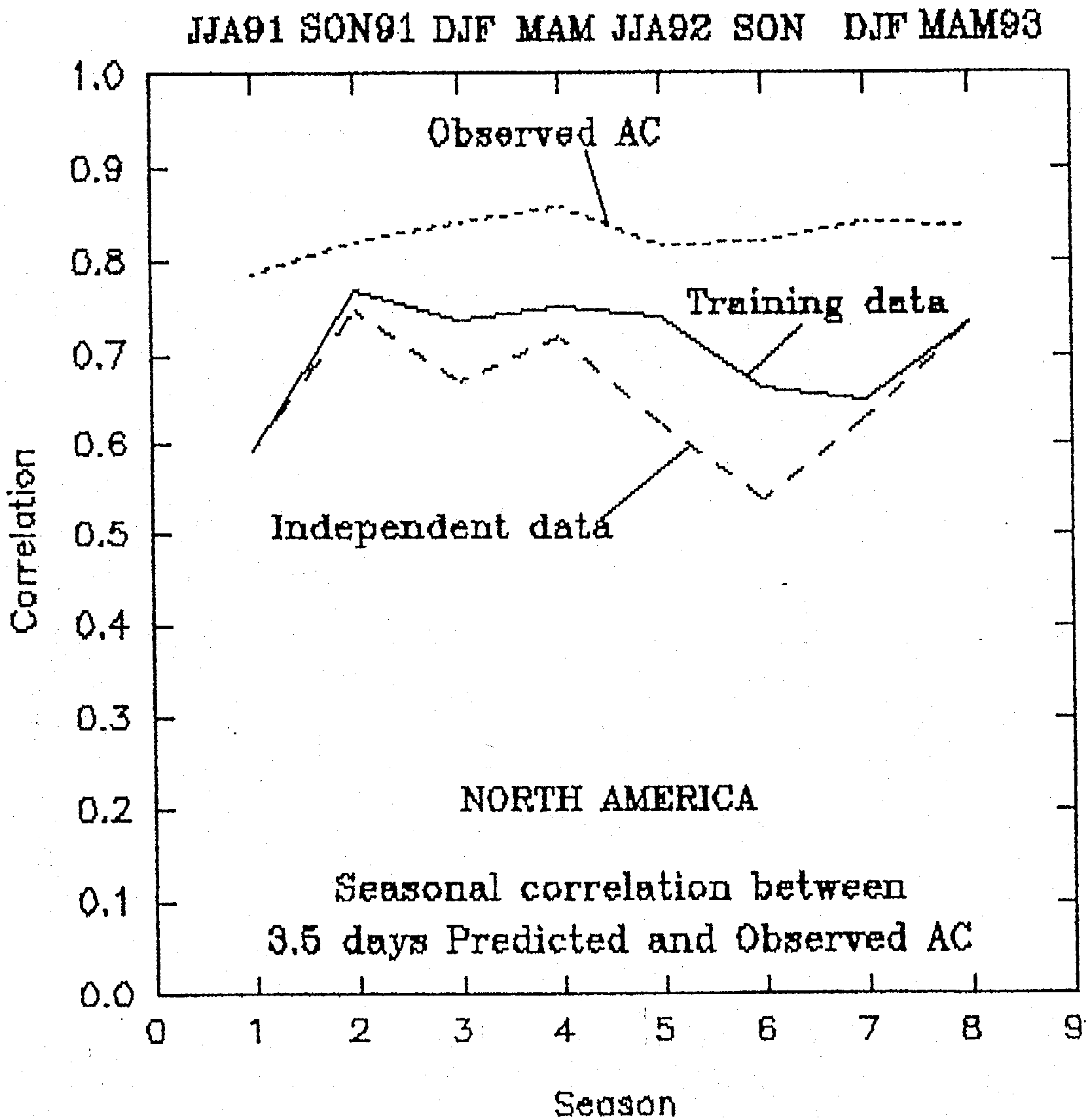


Fig.15a'

Fig. 15b

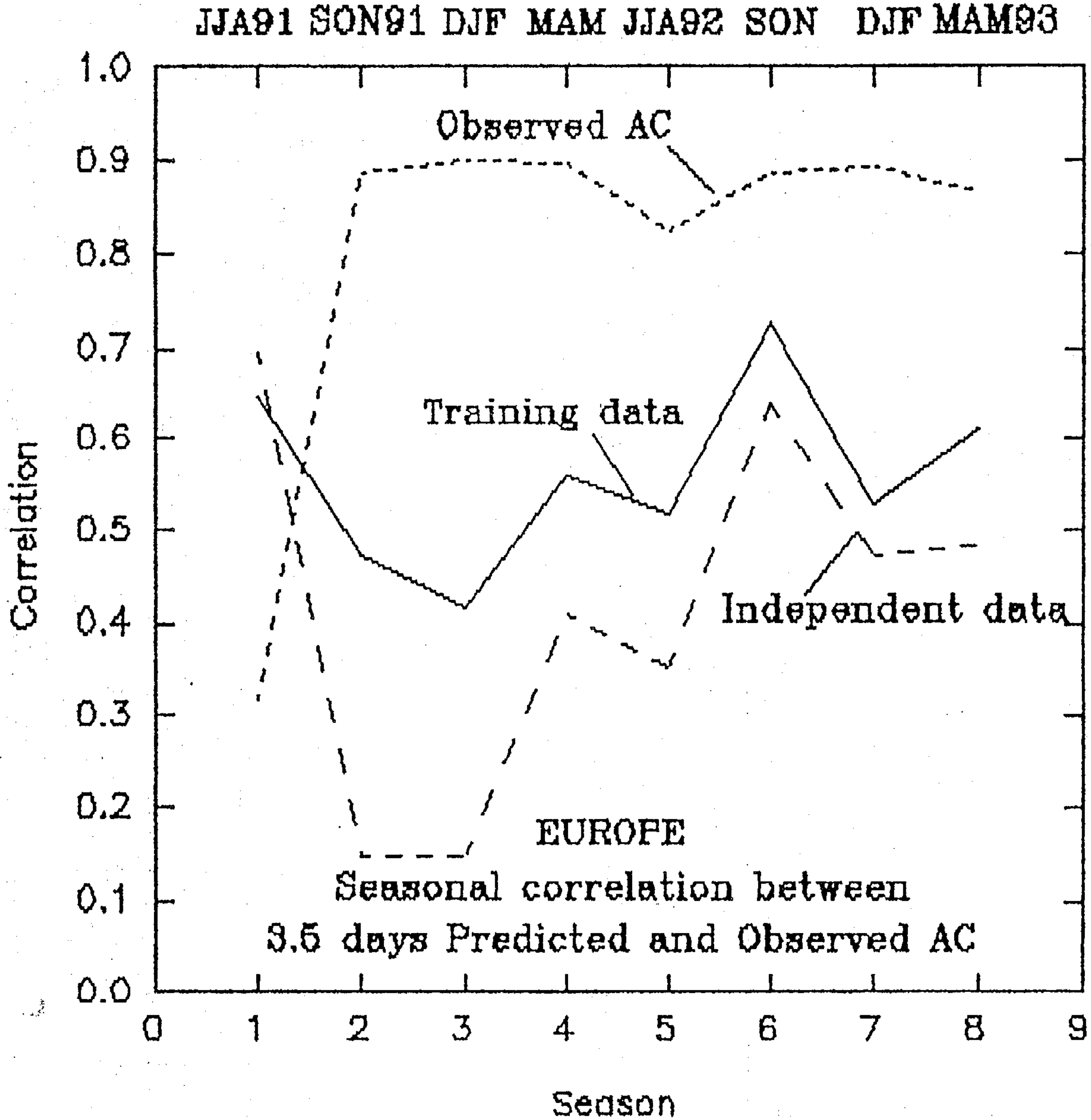


Fig 15b

Fig. 15c

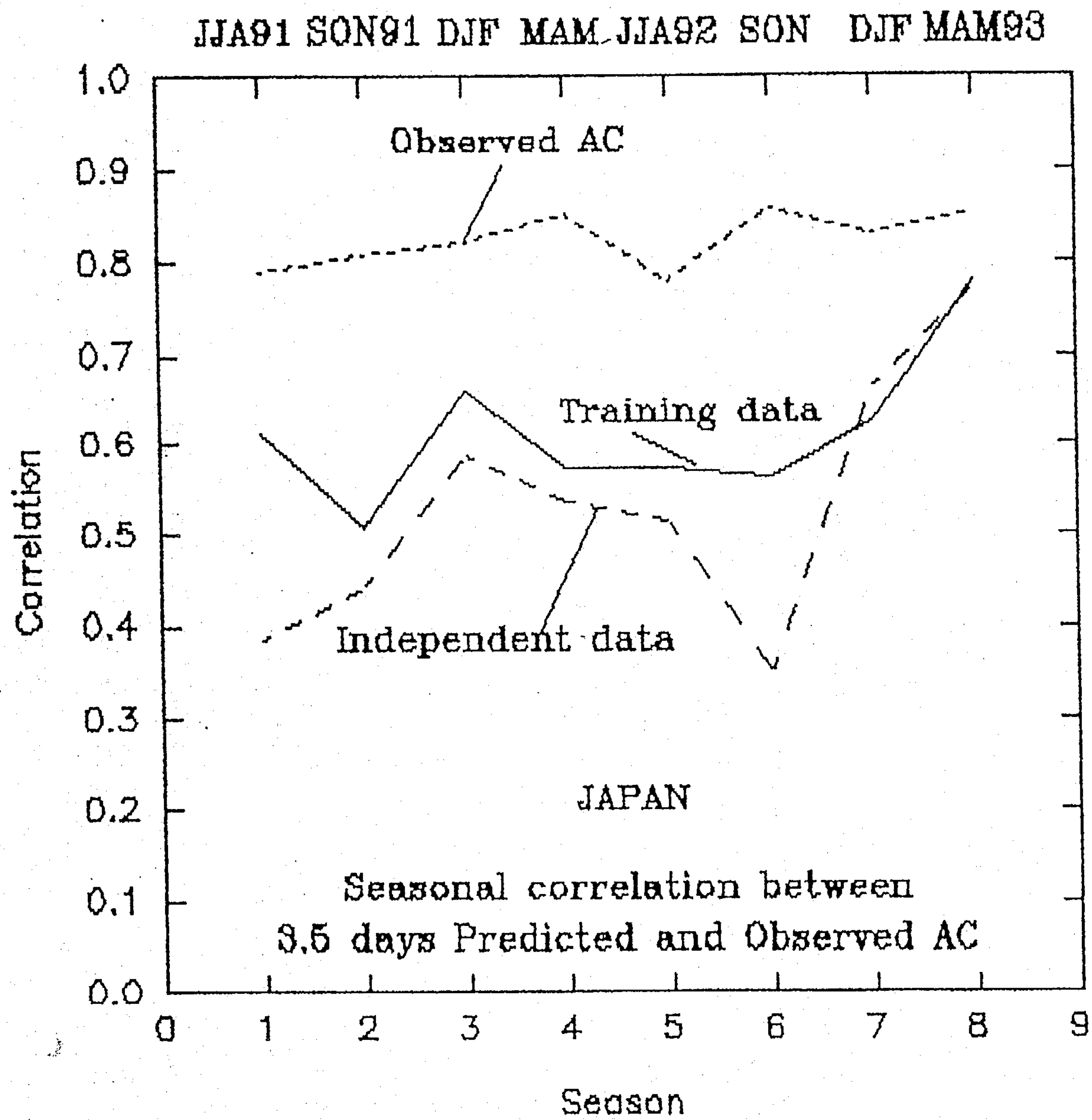


Fig. 15c

Fig. 15d

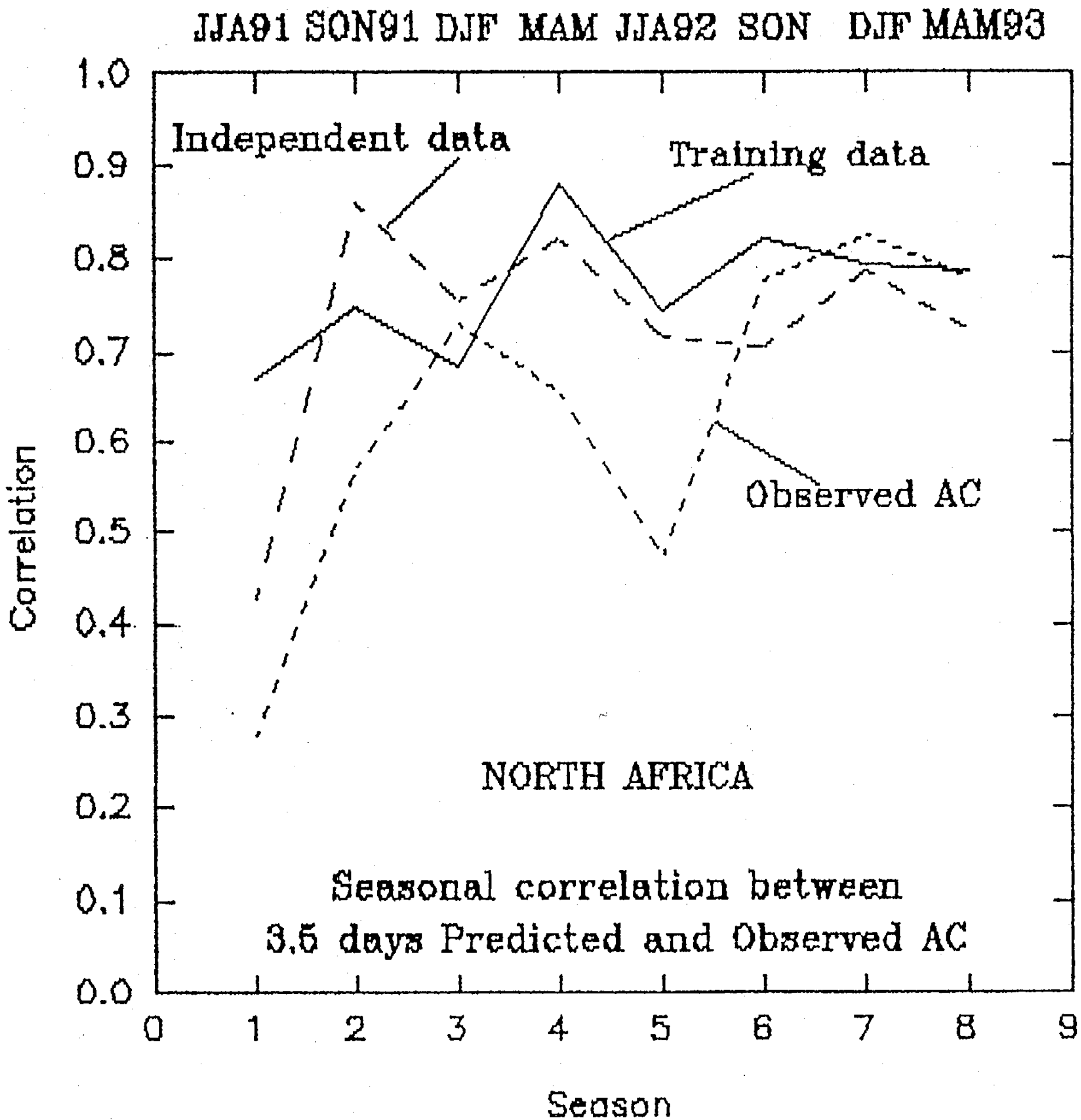


Fig. 15d